NORTH-HOLLAND
MATHEMATICS STUDIES 66

ANNALS
OF
DISCRETE
MATHEMATICS
16

Bonn Workshop on Combinatorial Optimization

Edited by
Achim Bachem
Martin Grötschel
Bernhard Korte

NORTH-HOLLAND

BONN WORKSHOP ON
COMBINATORIAL OPTIMIZATION

annals of discrete mathematics

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM • NEW YORK • OXFORD

NORTH-HOLLAND
MATHEMATICS STUDIES **66**

Annals of Discrete Mathematics (16)

General Editor: Peter L. Hammer

University of Waterloo, Ont., Canada

Bonn Workshop on Combinatorial Optimization

Based on lectures presented at the IV. Bonn Workshop on Combinatorial Optimization, August 28–30, 1980 organised by the Institute of Operations Research and sponsored by the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich 21

Edited by

Achim BACHEM
Martin GRÖTSCHEL
Bernhard KORTE

University of Bonn
W. Germany

1982

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM • NEW YORK • OXFORD

ISBN: 0 444 86366 4

Publishers
NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • NEW YORK • OXFORD

Sole distributors for the U.S.A. and Canada
ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE,
NEW YORK, N.Y. 10017

PRINTED IN THE NETHERLANDS

PREFACE

The field of combinatorial optimization has experienced a tremendous growth in recent years. This is for instance documented by the publication of many new scientific journals in this area as well as by the considerable number of large international conferences taking place every year.

Big meetings have the advantage of bringing a large number of people together and making a quick exchange of new results possible. Due to the (mostly) hectic atmosphere, however, they do not provide a platform for discussing problems in detail and digging deep into new aspects. This is the purpose of a workshop where few people gather together and even fewer people are given extensive time to present their ideas. Moreover, an informal atmosphere not restricted by time limits makes a more profound discussion of all aspects of the new developments possible.

From August 28 to August 30, 1980 the IV. Bonn Workshop on Combinatorial Optimization was held at the Rheinische Friedrich-Wilhelms-Universität, Bonn. It was organized by the Institut für Ökonometrie und Operations Research and generously sponsored by the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich 21.

Altogether 54 scientists from 16 different countries gathered at this meeting in a highly stimulating atmosphere. This volume constitutes a part of the outgrowth of the workshop and is based on the lectures presented there. The papers cover a broad spectrum of the field from submodular functions to perfect graphs, and from vertex packing to scheduling and subtree extension. All papers were subjected to a careful refereeing process.

We would like to express our sincere thanks to all authors for their cooperation, to all referees for their outstanding (albeit anonymous) contributions, and to the editor and publishers of this series for their support of this venture.

Bonn, October 1981

Achim BACHEM
Martin GRÖTSCHEL
Bernhard KORTE

CONTENTS

LIST OF PARTICIPANTS

J. Araoz, *Caracas, Venezuela*
G. Ausiello, *Roma, Italy*
A. Bachem, *Bonn, West Germany*
E. Balas, *Pittsburgh, PA, USA* and *Köln, West Germany*
F. Barahona, *Grenoble, France*
R. Bland, *Ithaca, NY, USA*
B. Bollobás, *Cambridge, England*
F.-J. Brandenburg, *Bonn, West Germany*
R.E. Burkard, *Köln, West Germany*
M. Burlet, *Grenoble, France*
L. Butz, *Bonn, West Germany*
G. Calvillo, *Mexico-City, Mexico*
G. Ciobanu, *Bucuresti, Romania*
G. Cornuéjols, *Pittsburgh, PA, USA*
R.W. Cottle, *Stanford, CA, USA*
U. Derigs, *Köln, West Germany*
R. Euler, *Köln, West Germany*
C. Fabian, *Bucuresti, Romania*
J. Fonlupt, *Grenoble, France*
E. Gabovich, *München, West Germany*
T. Gal, *Hagen, West Germany*
M. Gondran, *Clamart, France*
M. Grötschel, *Bonn, West Germany*
H. Hamacher, *Köln, West Germany*
P.L. Hammer, *Waterloo, Canada* and *Lausanne, Switzerland*
W.L. Hsu, *Louvain-la-Neuve, Belgium*
E.L. Johnson, *Yorktown Heights, NY, USA* and *Bonn, West Germany*
M. Jünger, *Bonn, West Germany*
V. Klee, *Seattle, Washington, USA* and *Bonn, West Germany*
B. Korte, *Bonn, West Germany*
E.L. Lawler, *Berkeley, CA, USA*
J.K. Lenstra, *Amsterdam, Holland*
T. Liebling, *Zürich, Switzerland*
L. Lovász, *Szeged, Hungary*
R. Möhring, *Aachen, West Germany*

D.J. Naddef, *Grenoble, France*
G.L. Nemhauser, *Ithaca, NY, USA*
M. Padberg, *New York, NY, USA* and *Le Chesnay, France*
M. Plummer, *Nashville, TN, USA* and *Bonn, West Germany*
A. Prodon, *Zürich, Switzerland*
W.R. Pulleyblank, *Grenoble, France* and *Bonn, West Germany*
R. von Randow, *Bonn, West Germany*
A.H.G. Rinnooy Kan, *Rotterdam, Holland*
D. Romero, *Mexico-City, Mexico*
R. Schrader, *Bonn, West Germany*
A. Schrijver, *Amsterdam, Holland*
L. Slominski, *Warsaw, Poland*
M. Sysło, *Wroclaw, Poland*
J. Tind, *Aarhus, Denmark*
J.P. Uhry, *Grenoble, France*
M. Vlach, *Prag, CSSR*
L. Wolsey, *Louvain-la-Neuve, Belgium*
K. Zimmermann, *Prag, CSSR*
U. Zimmermann, *Köln, West Germany*

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 1–26

PARITY GRAPHS

M. BURLET

U.S.M.G., BP 53 X, 38041 Grenoble Cedex, France

J.P. UHRY

Université Scientifique et Médicale CNRS, IMAG BP 53 X, 38041 Grenoble Cedex, France

A graph $G = (V, E)$ is a parity graph if and only if for every pair of vertices (x, y) of G all the minimal chains joining x and y have the same parity.

A characterization of these graphs can be given by a condition on the odd cycles: parity graphs are just the graphs in which every odd cycle has two crossing chords. A theorem of Sachs states that these graphs are perfect.

These graphs are then studied from the algorithmic viewpoint. Polynomial algorithms are defined to recognize them, and to solve the following problems: maximum independent set, minimum coloring, minimum covering by cliques, maximum clique.

1. Introduction

It is rather strange that, when a class of perfect graphs has been characterized, the algorithmic aspect is seldom studied. In particular, there are not only the classical problems of perfect graphs (finding a maximum stable set and a minimum coloring), but also the major problem of recognizing such a class of graphs in polynomial time.

These problems remain unsolved for many classes of perfect graphs (Meyniel's graphs [11], perfect 3-chromatic graphs [16], perfect planar graphs [15]). The only exception is the general paper of Grötschel, Lovász and Schrijver [7] which gives a polynomial algorithm for maximum weighted independent set and minimum coloring for all perfect graphs. This algorithm based on the ellipsoid method unfortunately gives no idea of the structure of perfect graphs, and at the present time appears to be of no great combinatorial interest.

There exist classes of perfect graphs for which these problems are solved: bipartite graphs and their line graphs, triangulated graphs, comparability graphs, and their complements. The recognition problem is also solved for a number of subclasses of these latter graphs (see [6]).

Finally there are classes of graphs for which these problems are not all solved (for example perfect claw-free graphs ([8, 9, 12]), for which the recognition is not yet settled, to our knowledge).

In this study, we deal with a particular class of perfect graphs which is a fairly natural extension of bipartite graphs: parity graphs.

Notation. Let $G = (V(G), E(G))$ be a graph with vertex set $V(G)$ and edge set $E(G)$. When no confusion is possible we will write V and E for $V(G)$ and $E(G)$.

Definition 1. A *minimal chain* is an elementary chain which is an induced subgraph.

In the graph of Fig. 1, chains (x, z, t, v, y) and (x, z, u, v, y) are minimal but not, for example, (x, z, t, u, v).

Definition 2. The *parity of a minimal chain* is the parity of the number of its edges.

In particular, if two vertices x and y are adjacent, the only minimal chain joining them is the chain reduced to the edge (x, y); this chain is odd.

Definition 3. A (simple, undirected) graph $G = (V, E)$ is called a *parity graph* if, for every pair of vertices x and y of G, all minimal chains joining x and y have the same parity.

Clearly, the notion of a parity graph generalizes that of a bipartite graph. Cliques are non-bipartite parity graphs. The graph depicted in Fig. 2 is a less trivial example.

In Section 2 we prove that a graph G is a parity graph, if and only if each odd cycle of length at least five contains two crossing chords. A theorem of Sachs [13] enables us to confirm that these graphs are perfect.

In Section 3 we prove some properties of these graphs, and we specify their minimal separating sets. These results can be compared to those of Gallai [5] for o-triangulated graphs.

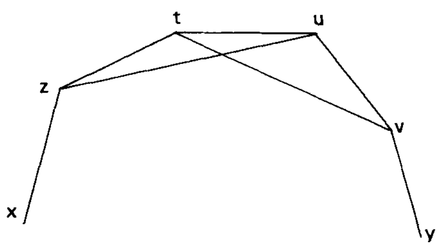


Fig. 1.

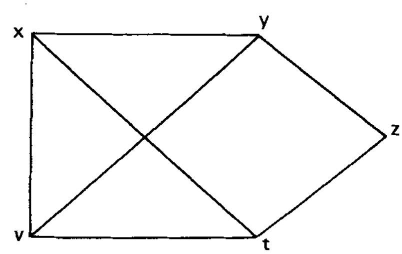


Fig. 2.

In Section 4 a polynomial algorithm for parity graph recognition is given. It is also shown that these graphs are in fact built from two classes of perfect graphs: bipartite graphs, and 'cographs' studied by Corneil, Lerchs and Stewart [3], here called 2-parity graphs.

More precisely: the class of perfect graphs is closed under making true or false twins (that is replacing a vertex by a set of two vertices linked or not by an edge) [10] and under certain extensions by bipartite graphs (cf. Definition 14). It will follow that parity graphs are exactly those graphs arising by these operations from a single point.

Finally, in Section 5, polynomial algorithms are defined for the four above-mentioned problems (in cardinality and in weight). This will yield another proof of the fact that these graphs are perfect.

2. Characterization

Definition 4. We say that *two chords* (x, y) and (z, t) of an elementary cycle *cross*, if the vertices x, z, y, t are different and in this order on the cycle.

Theorem 1. *A graph* $G = (V, E)$ *is a parity graph, if and only if every odd elementary cycle has two crossing chords.*

Proof. *Necessary condition*: The condition is necessary for a cycle of 5.

Suppose we admit the property on an odd cycle of cardinality k ($k > 5$) and prove it is still true for a cycle of cardinality $k + 2$.

It is easy to check that such an odd cycle contains at least two chords, and two chords which do not cross create at least one odd cycle, whose cardinality is lower than or equal to k, and the property follows by induction.

Sufficient condition: Take a graph G which verifies the condition and which is not a parity graph. As the structure we want for G is hereditary (under taking induced subgraphs) we shall choose a counter-example which is minimal with respect to the vertices.

This counter-example has two vertices x and y joined by an even minimal chain $(x, u_1, \ldots, u_t, y)$ and an odd minimal chain $(x, v_1, \ldots, v_s, y)$ with $s \geqslant 2$ (cf. Fig. 3).

Let

$$i_0 = \min\{i \mid \exists j > 1: (u_i, v_j) \in E\},$$

$$j_0 = \max\{j \mid (u_{i_0}, v_j) \in E\}.$$

So $i_0 < t$ and $j_0 > 1$, since the odd cycle formed by these two chains has two

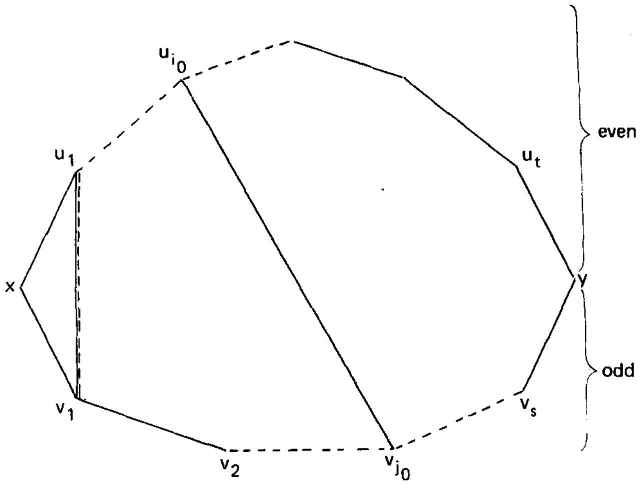


Fig. 3.

crossing chords by assumption. Now $(u_{i_0}, v_{j_0}, v_{j_0+1}, \ldots, v_s, y)$ and $(u_{i_0}, u_{i_0+1}, \ldots, u_t, y)$ are minimal chains, having the same parity, as they are contained in a smaller graph than G. Similarly, $(x, u_1, \ldots, u_{i_0}, v_{j_0})$ and $(x, v_1, \ldots, v_{j_0})$ have the same parity. However, the sum of the length of these four chains, $(s - j_0 + 2) + (t - i_0 + 1) + (i_0 + 1) + j_0 = s + t + 4$, is odd, which is a contradiction.

Theorem 2 (Sachs [13]). *Parity graphs are perfect.*

Without proof, we mention an obvious corollary.

Corollary 3. *A graph $G = (V, E)$ is a parity graph if and only if it does not contain any of the following configurations as induced subgraphs:*
- Δ_{2k+1} *odd cycle, without chord, on* $2k + 1$ *vertices,* $k \geqslant 2$ *(also called odd hole).*
- Δ'_{2k+1} *odd cycle, with only one short chord, on* $2k + 1$ *vertices,* $k \geqslant 2$ *(a short chord is a chord giving birth to a triangle).*
- $\tilde{\Delta}_5$ *cycle on* 5 *vertices with two non-crossing chords.*

Here, it seems interesting to recall two related results:

(a) the result of Gallai [5] and Surányi [14] that ‘o-triangulated graphs’ (in which every odd elementary cycle contains at least two uncrossing chords) are perfect, and

(b) the more general result of Meyniel [11] that each graph which has two chords in every odd elementary cycle is perfect.

In Fig. 4, we give an example of a graph which contains two chords in each of its odd cycles but which is, however, neither an o-triangulated graph nor a parity graph.

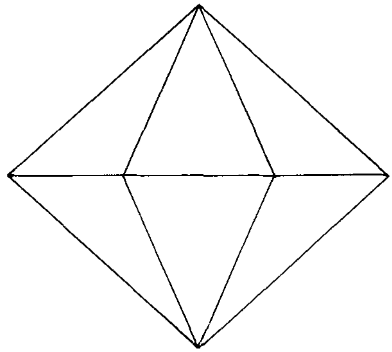


Fig. 4.

3. Description and properties

Notation. For a graph $G=(V,E)$ and $S\subseteq V$ we shall denote by $G(S)$ the subgraph induced by S. We denote by Γ_x the set of vertices adjacent to x, i.e.,

$$\Gamma_x=\{y\in V \mid (x,y)\in E\}.$$

For $A\subseteq V$ the intersection $\Gamma_x\cap A$ will be denoted by $\Gamma_x(A)$. Occasionally, when $H=G(A)$ is a subgraph of G we shall write $\Gamma_x(H)$ for $\Gamma_x(A)$.

Definition 5. We call two vertices x and y *true twins* if they are joined by an edge and have the same adjacents except for x and y (that is, $\Gamma_x\backslash\{y\}=\Gamma_y\backslash\{x\}$). Two vertices x and y are called *false twins* if they are not joined and have the same adjacents.

By Lovász [10] we know that the operation which consists of adding one (true or false) twin to a vertex of a perfect graph builds a new perfect graph.

In addition, this operation applied to a parity graph leaves a parity graph. This is false, however for o-triangulated graphs (Fig. 5), but true again for Meyniel graphs [11].

Definition 6. A graph without twins will be called *prime*.

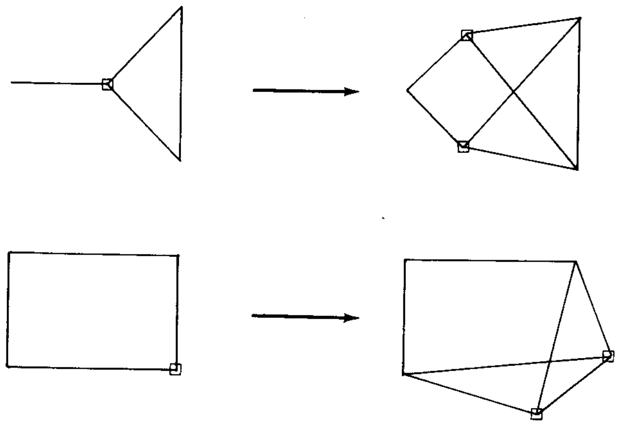


Fig. 5.

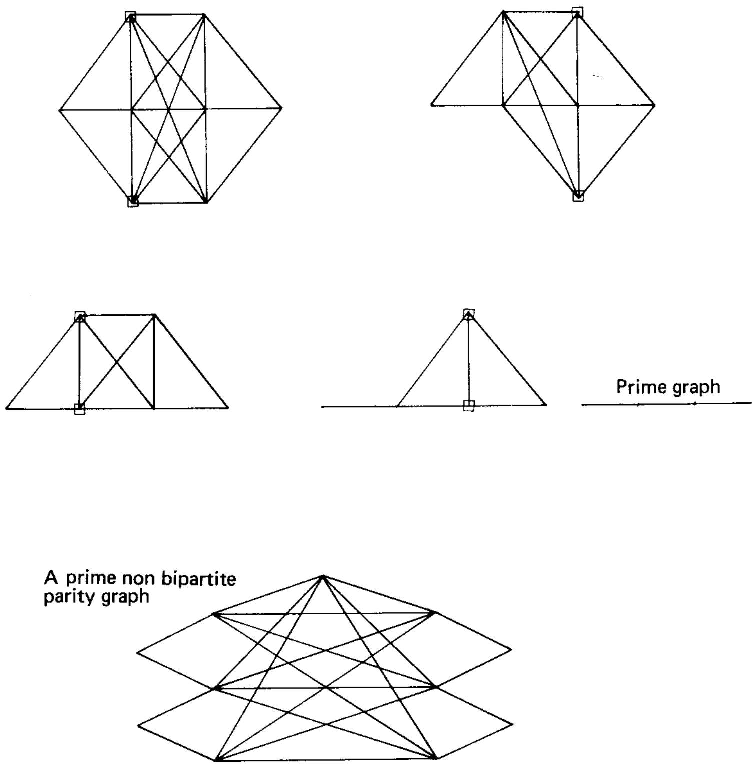


Fig. 6.

In Fig. 6 we give an example of the reduction of a parity graph, and an example of a prime parity graph.

Definition 7. In a parity graph $G = (V, E)$, *the partition induced by the vertex* x, denoted by (P_x, I_x), is the ordered partition of the vertices of G into the classes P_x and I_x, where P_x (resp. I_x) is the set of vertices of V joined to x by an even minimal chain (odd minimal chain, respectively). We assume that x is joined to x by an even minimal chain.

Notation. We may consider only the restriction of the bipartition induced by x to a subset A of V. Then we denote: $P_x(A) = P_x \cap A$ and $I_x(A) = I_x \cap A$.

Lemma 4. *Any minimal separating set A of a parity graph $G = (V, E)$ can be partitioned into two parts denoted R and B which have the following property: the vertices of R induce the same partition in $V \backslash A$ and the vertices of B the opposite partition in $V \backslash A$, that is,*

$$\forall r_1, \forall r_2 \in R \quad P_{r_1}(V\backslash A) = P_{r_2}(V\backslash A) \quad \textit{and hence} \quad I_{r_1}(V\backslash A) = I_{r_2}(V\backslash A),$$

$$\forall r_1 \in R, \forall b_1 \in B \quad P_{r_1}(V\backslash A) = I_{b_1}(V\backslash A) \quad \textit{and hence} \quad I_{r_1}(V\backslash A) = P_{b_1}(V\backslash A).$$

Proof. We suppose $|A| > 1$ (otherwise, it would be obvious). Let CX_1 and CX_2 be two different connected components of the subgraph induced on $V\backslash A$ (cf. Fig. 7).

Let x_1 belong to CX_1, x_2 belong to CX_2, z and t be two different vertices of A. A being minimal there exist minimal chains $C_1(x_1, x_2)$ and $C_2(x_1, x_2)$ joining x_1 and x_2, there only vertex from A being z for $C_1(x_1, x_2)$ and t for $C_2(x_1, x_2)$. Because $C_1(x_1, x_2)$ and $C_2(x_1, x_2)$ have the same parity we have $P_t(V\backslash A) = P_z(V\backslash A)$ or $P_t(V\backslash A) = I_z(V\backslash A)$.

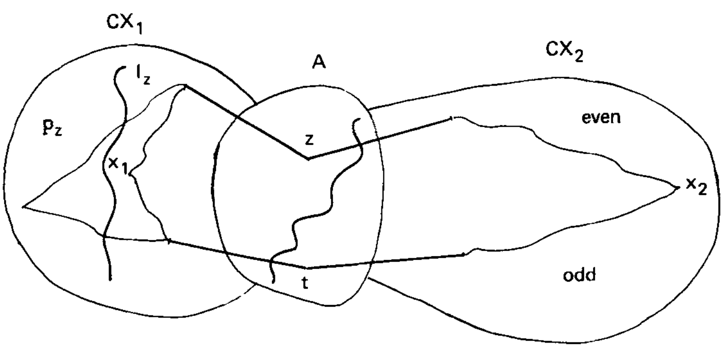


Fig. 7.

Lemma 5. *Let A be a minimal separating set partitioned into R and B as in Lemma 4. If $r_1, r_2 \in R$ and $(r_1, r_2) \in E$, then $\Gamma_{r_1}(V \backslash A) = \Gamma_{r_2}(V \backslash A)$.*

Proof. Suppose the property is false, and let $x \in \Gamma_{r_1}(V \backslash A)$ and $x \notin \Gamma_{r_2}(V \backslash A)$ (cf. Fig. 8). Then (r_2, r_1, x) is a minimal chain of even parity, and hence $x \in I_{r_1}$ and $x \in P_{r_2}$, contradicting Lemma 4.

Lemma 6. *With the same hypotheses as those of the preceding lemma, if $r_1, r_2 \in R$ and $(r_1, r_2) \in E$, then $\Gamma_{r_1}(B) = \Gamma_{r_2}(B)$.*

Proof. Suppose $b_1 \in \Gamma_{r_1}(B)$ and $b_1 \notin \Gamma_{r_2}(B)$, and let x be a vertex in $V \backslash A$ joined to b_1 (such a x exists by minimality of A) (cf. Fig. 9). As (r_2, b_1), (r_1, x) and (r_2, x) are not in E, there is a minimal chain from r_1 to x with length two, and another minimal chain from r_2 to x with length three. This contradicts Lemma 4.

Lemma 7. *With the same hypotheses as before, a connected component of the subgraph induced by the vertices of R has no minimal chain of length three.*

Proof. Suppose there is such a chain (r_1, r_2, r_3, r_4). Let us consider a vertex x of $V \backslash A$ which is adjacent to r_1. From Lemma 5 we know that x is also adjacent to r_2, r_3, and r_4. The subgraph induced by these vertices is a 5-cycle, chords of which may only come from x. This is a contradiction.

Remark 1. This property remains true if for the set R we take the adjacents of any one of the vertices of a parity graph. (They form a separating set which need not be minimal.)

Definition 8. A *2-parity graph* is a graph in which the length of all minimal chains is at most two.

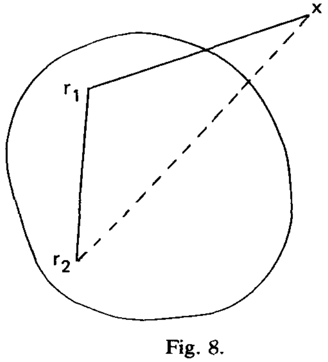


Fig. 8.

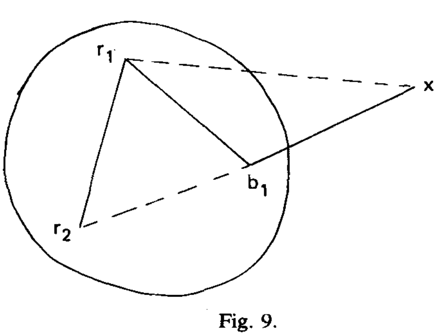


Fig. 9.

Examples of 2-parity graphs are cliques, and, more generally, complete multipartite graphs (a complete multipartite graph is a graph in which the vertices can be partitioned into stable sets, where two vertices are adjacent if they belong to different classes. Such graphs are o-triangulated [5].

Below we give a characteristic property of a 2-parity graph (for other properties, see [3]).

Lemma 8. *A connected 2-parity graph with more than one vertex has at least two (true, or false) twins.*

Proof. For a clique, it is obviously true. Otherwise, let x and y be two vertices which are not joined. There exists a minimal separating set A which separates x from y.

Let CX_x and CX_y be their respective connected components, in the subgraph induced by $V\backslash A$.

For each vertex z of CX_x, z is adjacent to all the vertices of A, otherwise there would be a minimal chain of length three.

When $CX_x = \{x\}$ and $CX_y = \{y\}$, then x and y are false twins, else at least one of CX_x and CX_y has cardinality greater than one, say $|CX_x| > 1$. The proof continues by induction in CX_x. Twins in subgraph CX_x will effectively be twins in the initial graph, because they have the same adjacents in A.

Corollary 9. *A graph is a 2-parity graph if and only if it arises from a single point by adding true or false twins* (*cf. Fig.* 10).

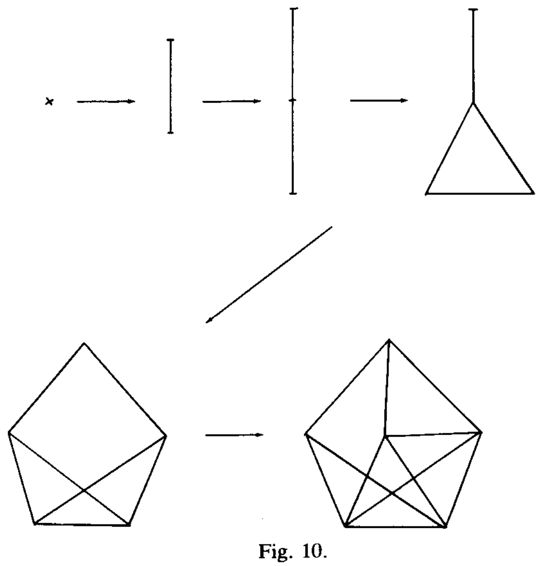


Fig. 10.

Theorem 10. *In a prime parity graph, every minimal separating set is a bipartite graph (one of the two parts can be empty).*

Proof. This theorem directly follows from the preceding lemmas, where we have shown that a connected component of the graph $G(R)$ (or $G(B)$) was a 2-parity graph and that for every two vertices r_1 and r_2 of such a component,

$$\Gamma_{r_1}(B) = \Gamma_{r_2}(B) \quad \text{and} \quad \Gamma_{r_1}(V \backslash A) = \Gamma_{r_2}(V \backslash A).$$

True or false twins, relating to this component are therefore twins in graph G. Thus, by removing true or false twins, we can reduce this component to only one vertex.

One further property of parity graphs is given below without proof.

Lemma 11. *In a two-connected parity graph, if two joined vertices x and y have disjoint neighbourhoods, then $I_x = P_y$.*

4. A constructive polynomial algorithm for recognition

To recognize a parity graph, we shall 'hang up' the graph on a vertex, and study the structure of the different levels (defined below). It will lead us not only to the polynomial algorithm to recognize parity graphs, but also to a theorem which gives a very easy constructing characterization of parity graphs.

This is the basis of the optimization algorithms described in Section 5. We shall study here the partition (P_a, I_a) induced by a well-defined vertex a of a parity graph $G = (V, E)$.

Definition 9. We shall note $C(x, y)$ a *shortest chain* (in number of edges) joining x to y.

Definition 10. We call the set of vertices of G at a distance i from a *the level i*, denoted by N_i, that is,

$$N_i = \{x \in V \mid |C(a, x)| = i\}.$$

The vertices of N_i are linked to a by minimal chains of the parity of i, so N_0 only contains vertex a and N_1 is the set of neighbourhoods of a. We will note by m the cardinality of the longest minimal chain starting at a and we will set $N_{m+1} = \emptyset$.

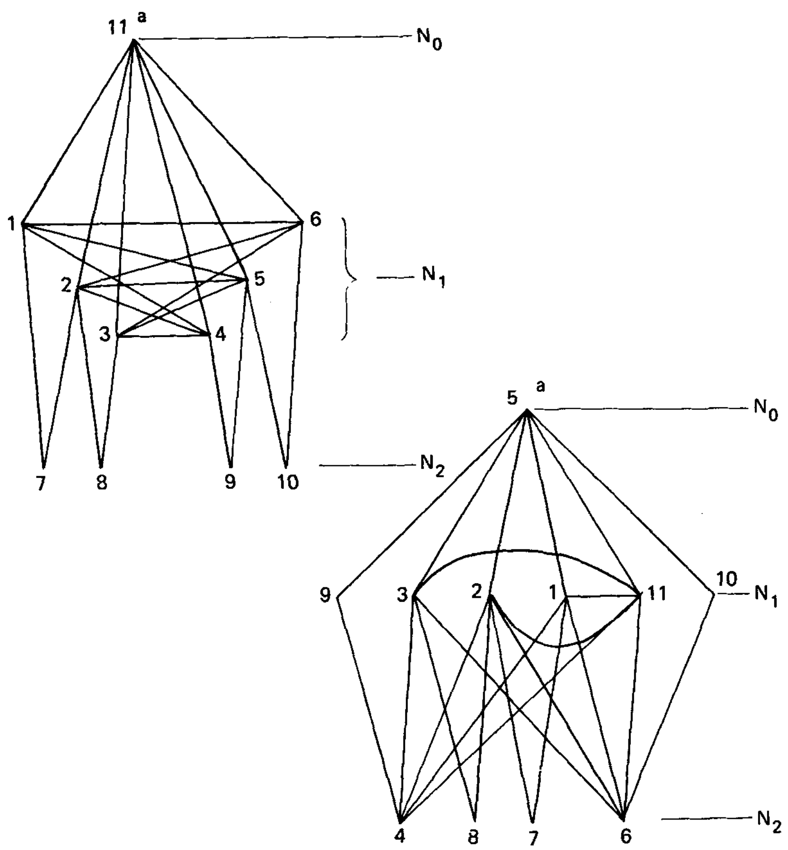


Fig. 11.

Remark 2. Dijkstra's shortest path algorithm enables us to find the partition of V into levels in polynomial time.

Remark 3. An edge $(x, y) \in E$ has its end points either on the same level or on two successive levels.

In Fig. 11 we give two possible 'hangings' of the graph of Fig. 6.

Lemma 12. *All vertices of a connected component of the subgraph $G(N_i)$ have the same adjacents at the level N_{i-1}.*

Proof. Let $x, y \in N_i$ and $(x, y) \in E$, let $x' \in \Gamma_x(N_{i-1})$ (cf. Fig. 12), and suppose that $(y, x') \notin E$.

As the edge (x', y) does not exist, by concatenating a chain $C(a, x')$ to the

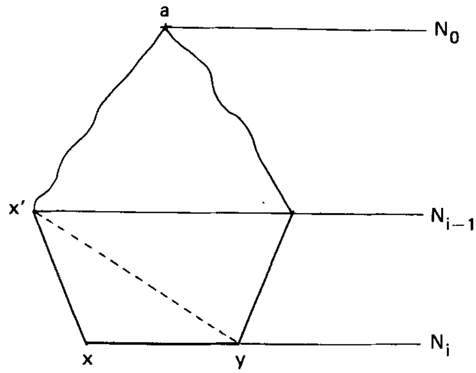


Fig. 12.

chain (x', x, y), it is possible to find a chain joining a to y which has the same parity as $i+1$. As there exists, by definition, one chain from a to y having the same parity as i, we have a contradiction.

This lemma shows that the subgraph induced by a level is a 2-parity graph (cf. Lemma 7).

Definition 11. We remove the level N_i, and we denote by $G_i^1, G_i^2, \ldots, G_i^q$ the different connected components obtained in the subgraph induced by the lower levels,

$$N_{i+1} \cup N_{i+2} \cup \cdots \cup N_m .$$

We denote by X_i^k the vertices of N_i adjacent to the component G_i^k.

Remark 4. The sets X_i^k are not necessarily disjoint and do not necessarily cover all the vertices of N_i.

Lemma 13. *If $x, y \in X_i^k$ and $(x, y) \in E$, then $\Gamma_x(G_i^k) = \Gamma_y(G_i^k)$.*

Proof. If the property is false, there exists a configuration illustrated by Fig. 13, where $(x, u) \notin E$.

Let $u \in \Gamma_y(N_{i+1})$, $v \in \Gamma_x(N_{i+1})$ and $(x, u) \notin E$.

- There exists a chain $C(a, u)$ of the parity of $i+1$.
- (u, y, x) is a minimal chain of length two, this means that we can find a minimal even chain $(u, \ldots, v, x)$ having all its vertices in G_i^k. Linking this chain to a chain $C(a, x)$, we obtain a chain of parity of i which connects u to a. This is a contradiction.

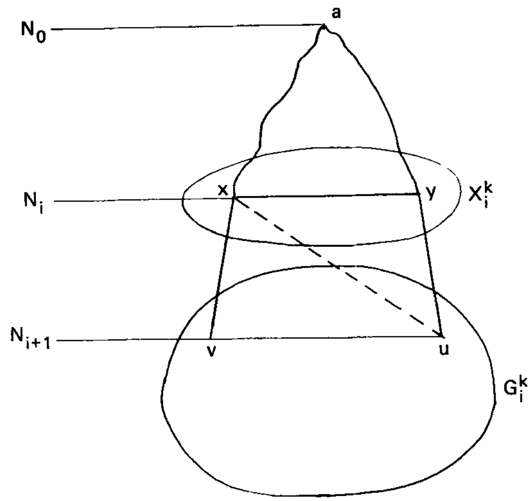


Fig. 13.

Lemma 14. *If $x, y \in X_i^k$, then $\Gamma_x(N_i \backslash X_i^k) = \Gamma_y(N_i \backslash X_i^k)$.*

Proof. Let $x, y \in X_i^k$ and $(y, z) \in E$. Let $z \in \Gamma_y(N_i \backslash X_i^k)$, and let $(x, v_1, \ldots, v_t, y)$ be the shortest path from x to y with vertices $v_1, \ldots, v_t$ in G_i^k, using at least one vertex of G_i^k. Let $(x, u_1, \ldots, u_s, y)$ be the shortest path from x to y with vertices $u_1, \ldots, u_s$ in $N_0 \cup \cdots \cup N_{i-1}$, using at least one vertex in $N_0 \cup \cdots \cup N_{i-1}$ (cf. Fig. 14). Then $(u_s, z) \in E$ (cf. Lemma 12). Consider the cycle

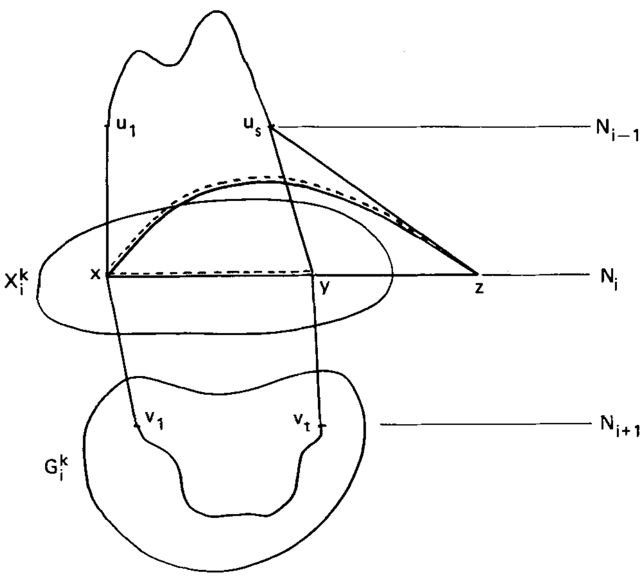


Fig. 14.

$(x, u_1, \ldots, u_s, z, y, v_t, \ldots, v_1, x)$. The only chords are (u_s, y) and, possibly, (x, y) and (x, z). Now, if (x, z) is not in E, the cycle will contain an odd cycle without two non crossing chords.

Definition 12. Associate at a given level N_i a graph $G' = (V', E')$. The vertices of G' are in one-to-one correspondence with the X_i^k. Two vertices of G' are *connected* by an edge of E' if the corresponding X_i^k and $X_i^{k'}$ have a non empty intersection, and are not included one in the other, that is,

$$X_i^k \cap X_i^{k'} \neq \emptyset, \qquad X_i^k \backslash X_i^{k'} \neq \emptyset, \qquad X_i^{k'} \backslash X_i^k \neq \emptyset.$$

Let $C_1, \ldots, C_p$ be the connected components of G'. Let $U_i^t = \bigcup_k (X_i^k)$, where k ranges over the vertices in C_t. It is easy to check that the family $\mathcal{U}_i = (U_i^1, \ldots, U_i^p)$ is a nested family, partially ordered by inclusion. The family $\mathcal{U}_m$ is empty.

Lemma 15. *If X_i^k and $X_i^{k'}$ belong to the same component of G', then there is no edge (z, w) with $z \in X_i^k \backslash X_i^{k'}$, $w \in X_i^{k'}$.*

Proof. Suppose there is an edge (z, w) such that $z \in X_i^k \backslash X_i^{k'}$, $w \in X_i^{k'}$, where X_i^k and $X_i^{k'}$ belong to the same component of G', and suppose we have chosen z, w, k, k' in such a way that the distance d between k and k' in the graph G' is as small as possible. Let $k = k_0, \ldots, k_d = k'$ be the shortest path from k to k' in G'. Then $X_i^{k_0} \cap X_i^{k_1} \neq \emptyset$, say $z' \in X_i^{k_0} \cap X_i^{k_1}$.

If $d \geqslant 2$, by Lemma 14 also (z', w) is in E, and moreover $z' \notin X_i^{k'}$. So replacing z, w, k, k' by z', w, k_1, k' gives a lower distance in G', which is a contradiction. Hence $d = 1$.

Let $(z, v_1, \ldots, v_s, z')$ and $(z', u_1, \ldots, u_s, w)$ be shortest paths with internal points in G_i^k and $G_i^{k'}$, respectively, using at least one point of G_i^k and $G_i^{k'}$, respectively. One easily checks that s and s' are odd. Then $(z, v_1, \ldots, v_s, z', u_1, \ldots, u_s, w, z)$ is an odd cycle without crossing chords, which is a contradiction.

Definition 13. An element U_i^t of the family $\mathcal{U}_i$ is called *e-maximal* relatively to the edge (x, y) if $x \in U_i^t$, $y \notin U_i^t$ and U_i^t is maximal in $\mathcal{U}_i$ for this property.

Lemma 16. *In a parity graph, an e-maximal U_i^t not contained in any other e-maximal set, can be reduced by removal of true or false twins to a stable set. After these operations the vertices of U_i^t are false twins in the subgraph induced by $N_0 \cup \cdots \cup N_i \cup \bigcup_k V(G_i^k)$, where the union ranges over all k with $X_i^k \not\subseteq U_i^t$.*

Proof. Let U_i^l be a minimal e-maximal element of $\mathcal{U}_i$ and let $K = \{k \mid X_i^k \subseteq U_i^l\}$. Let (x, y) be an edge of $G(N_i)$ associated to U_i^l chosen to satisfy with U_i^l, Definition 13.

Let X_i^k and $X_i^{k'}$ where $k \in K$ and $k' \in K$, then by choice of U_i^l there is no edge (u, v) with $u \in X_i^k \backslash X_i^{k'}$ and $v \in X_i^{k'}$. Hence by Lemma 13 for any two vertices u and v of a component of $G(U_i^l)$ we have

$$\Gamma_u\Big(\bigcup_k V(G_i^k)\Big) = \Gamma_v\Big(\bigcup_k V(G_i^k)\Big) \quad \text{where } k \text{ ranges over } K.$$

For any two vertices u and v of $G(U_i^l)$ we have $\Gamma_u(N_i \backslash U_i^l) = \Gamma_v(N_i \backslash U_i^l)$ (cf. Lemma 14). This implies that $y \in \Gamma_u$ and $y \in \Gamma_v$, hence, by Lemma 12, $\Gamma_u(N_{i-1}) = \Gamma_v(N_{i-1})$.

Let us prove that for any two vertices u and v of $G(U_i^l)$ we have $\Gamma_u(\bigcup_k V(G_i^k)) = \Gamma_v(\bigcup_k V(G_i^k))$ where k ranges over all k with $X_i^k \not\subseteq U_i^l$.

Suppose there exists $k' \notin K$ with $w \in G_i^{k'}$, $(u, w) \in E$ and $(v, w) \notin E$. Then $u \in X_i^{k'}$ and hence $U_i^l \subseteq X_i^{k'}$. As U_i^l is e-maximal, by Lemma 15 we have $y \in X_i^{k'}$. Hence, by Lemma 13, $\Gamma_u(G_i^{k'}) = \Gamma_y(G_i^{k'}) = \Gamma_v(G_i^{k'})$, which implies that (v, w) is an edge, contradicting our assumption.

Each component of $G(U_i^l)$ is a 2-parity graph which can be reduced by removal of true and false twins to a unique vertex (cf. Lemma 8). The former material enables us to report the reduction inside a component of $G(U_i^l)$ on the whole graph $G = (V, E)$.

The previous lemma is the fundamental argument of the recognition algorithm given later on, which will allow us to demonstrate Theorem 17 below.

Definition 14. Let G^* be a graph, and let $B = (V_1 \cup V_2, A)$ be a bipartite graph with colour classes V_1 and V_2. Let V' be a collection of false twins in G^* (possibly $|V'| = 1$). Then the graph arising by identifying the vertices in V' with certain vertices in V_1 is called an *extension* of G^* by B (cf. Fig. 15).

Theorem 17. *Every connected parity graph $G = (V, E)$ is obtained from a single vertex by the following operations:*
- *Φ_1 creation of a false twin,*
- *Φ_2 creation of a true twin,*
- *Φ_3 extension by a bipartite graph,*

applied successively and in any order.

It is obvious that a graph obtained by the three operations Φ_1, Φ_2, Φ_3 as indicated by Theorem 17, is a parity graph.

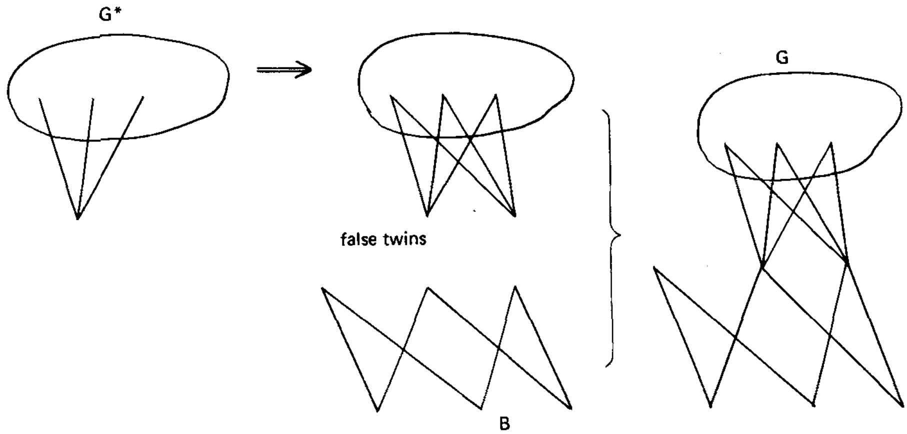


Fig. 15.

The following algorithm will in fact check that a graph which satisfies Lemmas 12, 13, 15 and 16 can be constructed by these three operations beginning with a single vertex.

The algorithm consists in looking for bipartite graphs which eventually extend (as in Φ_3) a graph hung by a vertex a. We carry out this research by beginning from the lower level N_m and then climbing from level to level. As soon as we detect that one of these bipartite graphs extends the remainder of the graph, we delete this bipartite graph. It is the inverse operation of Φ_3.

More precisely beginning by the level N_m, we are going to reduce into a stable set successively every level by the inverse of Φ_1, Φ_2 and Φ_3.

It is clear that the level N_m can be reduced to an independent set by the sole operations of reduction of a 2-parity graph (Φ_1^{-1} and Φ_2^{-1}).

Lemma 18. *If every level N_j $(j > i)$ is an independent set, then for all k the graphs G_i^k are bipartite graphs.*

Proof. Obvious.

Let us imagine that the levels N_j $(j > i)$ are reduced into independent sets. We describe the iteration which consists in reducing N_i into a stable set. In this iteration we are mainly concerned with the e-maximal U_i^t of $\mathcal{U}_i$.

For each one, starting with the minimal ones and after verification that it satisfies Lemma 16, we delete from G by Φ_3^{-1} a bipartite graph. This enables us to reduce $G(U_i^t)$ by Φ_2^{-1} and Φ_1^{-1} to a single vertex. When the list of e-maximal elements of $\mathcal{U}_i$ is exhausted, we can reduce by Φ_1^{-1} and Φ_2^{-1} each connected component of $G(N_i)$ to a single vertex.

Definition 15. In the same way as the G_i^k corresponds to a X_i^k we will associate a graph H_i^t to an element U_i^t of the family $\mathcal{U}_i$ with

$$V(H_i^t) = \bigcup_k V(G_i^k) \quad \text{and} \quad E(H_i^t) = \bigcup_k E(G_i^k)$$

where k ranges over the vertices in C_t.

Recognition Algorithm for a Parity Graph

begin
For $i := m$ downto 1 do
begin
Build the ordered family $\mathcal{U}_i$, find all the e-maximal elements of $\mathcal{U}_i$.
while $\mathcal{U}_i \neq \emptyset$ do
  begin
  choose U_i^t a minimal element of $\mathcal{U}_i$.
  Delete U_i^t from $\mathcal{U}_i$.
  If U_i^t is not e-maximal nor a maximal element of $\mathcal{U}_i$ then
    begin
    let $U_i^{t'}$ be an element of $\mathcal{U}_i$ containing U_i^t and minimal for this property
    $V(H_i^{t'}) := V(H_i^{t'}) \cup V(H_i^t)$; $E(H_i^{t'}) := E(H_i^{t'}) \cup E(H_i^t)$.
    end
  if U_i^t is e-maximal then
    begin
    Check that the vertices of $G(U_i^t)$ have the same neighbours in the subgraph defined in Lemma 16.
    Check that the vertices of a component of $G(U_i^t)$ have the same neighbours in H_i^t.
    Check that every connected component of $G(U_i^t)$ is a 2-parity graph and shrink it into a single vertex.
    Delete H_i^t by Φ_3^{-1}.
    Shrink $G(U_i^t)$ replacing it by a single vertex.
    end
  end
Check that each connected component of $G(N_i)$ has the same neighbours in N_{i+1} and N_{i-1}.
Also check that these components are 2-parity graphs and
Shrink them to a single vertex.
end
end

Note that every shrinking performed in this algorithm can be done by iterating Φ_1^{-1} and Φ_2^{-1}.

It is clear that this algorithm stops after a finite number of iterations, and it reduces any parity graph to a single vertex. So we have just established that a parity graph can be constructed following the operations indicated in Theorem 17.

Remark 5. The running time of the algorithm (Dijkstra's algorithm, building the ordered family $\mathcal{U}_i$, detecting the e-maximal elements of $\mathcal{U}_i$, checking for neighbours, recognizing that a graph is a 2-parity graph, ...) is bounded by the square of the number of vertices concerned by the operation (in [3] an algorithm is indicated to recognize a 2-parity graph whose running time is $O(n^2)$). The number of X_i^k (for all i and k) is less than $|V|$. So the number of vertices created by shrinking the e-maximal U_i^j is less than $|V|$. Hence the number of vertices examined by each operation is less than $2|V|$. The complexity of the algorithm presented is $O(|V|^2)$.

Remark 6. We have just demonstrated that a graph is a parity graph if and only if it verifies Lemmas 12, 13 and 16. If the graph is not a parity graph, the algorithm finds the lemma in fault. The proof of the lemma in question then allows us to detect an odd cycle which does not have two crossing chords. We may use this algorithm to exhibit in any non-parity graph, an odd cycle which does not verify the hypothesis of Theorem 1.

Remark 7. A new problem, apparently similar to that of recognizing a parity graph, is as follows:

> Let G be a graph, and let x and y be vertices of G. "Are all minimal chains joining x to y even?"

A polynomial algorithm for this problem would yield a polynomial algorithm to verify that a graph G has no chordless odd cycle. Indeed, enumerating all minimal chains of length 2 is polynomial. Let (x, z, y) be such a chain. Remove the vertex z, and the vertices $\Gamma_z - \{x\} - \{y\}$.

If all the minimal chains now joining x to y are even, then x, z, y do not belong to an odd chordless cycle, otherwise x, z, y do belong to an odd chordless cycle. The same algorithm in the complementary of G, would allow us to check the hypothesis of the strong perfect graph conjecture. This problem appears much more difficult.

We give here a theorem which is suggested by the preceding result. This theorem gives a direct proof of Sach's theorem (when Theorem 17 is known).

Theorem 19. *If we extend a perfect graph by a bipartite graph, we obtain a new perfect graph.*

Proof. It suffices to show that the cardinality of a maximum clique is equal to the chromatic number in all perfect graphs extended by a bipartite graph. Operation Φ_3 does not increase the clique or coloring number of a graph, except when the graph is trivial.

5. Polynomial algorithms in parity graphs

The interest of the characterization in the previous section is that it enables the construction of polynomial algorithms for the four problems of perfect graphs.

These algorithms differ from the more general algorithms as those proposed in [7] in that the present ones are based on the algorithm of finding a maximum weighted independent set in a bipartite graph, as described by Ford and Fulkerson [4]. This is also the case, for example, for the algorithms proposed in the latter book for comparability graphs. Here we aim at solving the problems for nonnegative integral weight functions c defined on V.

We might think in fact that it is sufficient to solve the problem of maximum cardinality, as the transformation of a vertex into false or true twins enables a problem with an integer valued cost function to be transformed into a cardinality problem. This transformation constructs a parity graph from a parity graph. Unfortunately, this transformation is not polynomial, and consequently, we shall study the more general optimization problems.

Let K be the matrix of all maximal cliques of $G=(V,E)$, where $|V|=n$, that is, each line of K is the characteristic vector of a maximal clique of G in $\{0,1\}^n$. Similarly, S is the matrix of all maximal stable sets of G.

The four problems paired by duality are as follows:

$$\begin{cases} \max z = \sum_i c_i x_i\,, \\ \text{subject to } Kx \leqslant 1\,, \\ \qquad x_i \geqslant 0\,, \end{cases} \qquad \begin{cases} \min w = \sum_i y_i\,, \\ \text{subject to } yk \geqslant c\,, \\ \qquad y_i \geqslant 0\,, \end{cases}$$

$$\begin{cases} \max z = \sum_i c_i x_i\,, \\ \text{subject to } Sx \leqslant 1\,, \\ \qquad x_i \geqslant 0\,, \end{cases} \qquad \begin{cases} \min w = \sum_i y_i\,, \\ \text{subject to } yS \geqslant c\,, \\ \qquad y_i \geqslant 0\,. \end{cases}$$

We solve each of the following two pairs simultaneously:

(i) Maximum weighted stable set and minimum covering by cliques, and

(ii) Maximum weighted clique and minimum covering by stable sets.

The method used, which is classical since the work of Edmonds on matching, (cf. for example, [2]), consists in finding for any nonnegative integral weight function c, defined on V, a pair of dual integral solutions x and y which satisfy the complementary slackness condition.

For the three operations Φ_1, Φ_2 and Φ_3 which transform G^* into graph G, we are content here to give the transformation on the weight. Starting from a dual optimal integral pair x^*, y^* defined on G^*, we shall indicate how to obtain a dual optimal integral pair x, y defined on G.

Notation. We shall denote by α the vertex of G^* concerned by the operations Φ_1, Φ_2, Φ_3 by c_α its weight, and by $K_1^\alpha, K_2^\alpha, \ldots, K_l^\alpha$ all the maximal cliques of G^* containing the vertex α, associated with strictly positive components of y^*, denoted respectively by

$$y^*(K_1^\alpha), y^*(K_2^\alpha), \ldots, y^*(K_l^\alpha).$$

These l cliques are the restrictions of $2l$ maximal cliques of the graph G in the case where Φ_1 transforms α in false twins a and b, denoted respectively by

$$K_1^a, K_2^a, \ldots, K_l^a, K_1^b, K_2^b, \ldots, K_l^b.$$

They are the restrictions of l maximal cliques in the case where Φ_2 transforms α in true twins a and b, denoted respectively by

$$K_1^{ab}, K_2^{ab}, \ldots, K_l^{ab}.$$

5.1. Maximum weighted independent set—minimum covering by maximal clique

5.1.1. False twins a and b

Particular case of 5.1.3 where

$$B = (V_1 \cup V_2, A), \qquad V_1 = (\{a\}, \{b\}), \qquad V_2 = \emptyset, \qquad A = \emptyset.$$

5.1.2. True twins a and b (cf. Fig. 16)

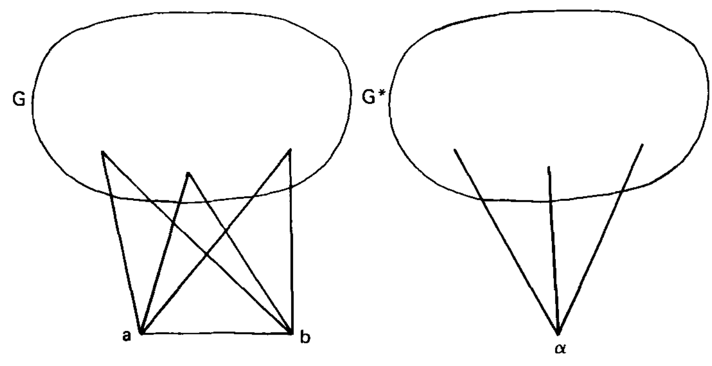


Fig. 16.

We define $c_\alpha = \max(c_a, c_b) = c_a$.

Transformation on the pair of optimal solutions

(i) Case where $x^*(\alpha) = 1$:

$$x(a) = 1\,, \qquad x(b) = 0\,,$$
$$y(K_1^{ab}) = y^*(K_1^{\alpha}), \ldots, y(K_l^{ab}) = y^*(K_l^{\alpha})\,;$$

(ii) Case where $x^*(\alpha) = 0$:

$$x(a) = 0\,, \qquad x(b) = 0\,,$$
$$y(K_1^{ab}) = y^*(K_1^{\alpha}), \ldots, y(K_l^{ab}) = y^*(K_l^{\alpha})\,.$$

The components of x and y which are not defined are unchanged. This pair x and y fulfills the complementary slackness conditions.

5.1.3. Extension by a bipartite graph

Without loss of generality we suppose that the bipartite graph $B = (V_1 \cup V_2, A)$ is only fixed by two false twins: a and b (cf. Fig. 17).

We define $c_\alpha = s^B_{\max} - s^B_{\bar a \bar b}$, where $s^B_{\max}$ stands for the value of a maximum weighted stable set in B and $s^B_{\bar a \bar b}$ stands for the value of a maximum weight stable set in B where vertex a and vertex b have been eliminated. (In general, we shall note the sets in capital letters, and the value in lower case letters.)

The idea of this transformation is quite simple. As soon as a vertex in the neighbourhood $\Gamma_\alpha(G^*)$ of a or b is in the maximum weight independent set of

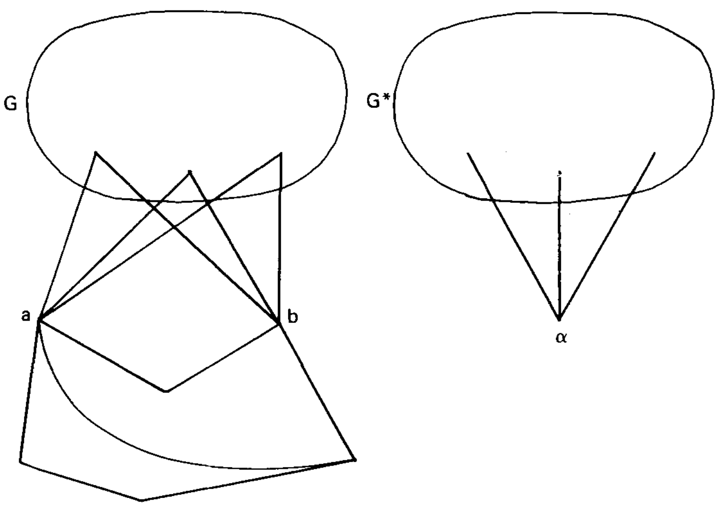


Fig. 17.

G it contains $S^B_{a\bar{b}}$. In the other case, it contains $S^B_{\max}$. We translate therefore by c_α the regret which results from putting α in the maximum weighted stable set of G^*.

Transformation on the pair of optimal solutions

(i) Case where $x^*(\alpha) = 1$. We obtain x by withdrawing from x^* its component $x^*(\alpha)$, and by extending this vector by a characteristic vector of $S^B_{\max}$.

(ii) Case where $x^*(\alpha) = 0$. We complete in this case with a representative vector of $S^B_{\bar{a}b}$.

To obtain y, the process is slightly more elaborate.

It is necessary to find an integer solution to the linear system:

$$\begin{cases} y(K_1^a) + y(K_1^b) = y^*(K_1^a), \\ y(K_l^a) + y(K_l^b) = y^*(K_l^a), \\ y(K_1^a) + y(K_2^a) + \cdots + y(K_l^a) = \xi_a, \\ y(K_1^b) + y(K_2^b) + \cdots + y(K_l^b) = \xi_b. \end{cases} \tag{1}$$

(i) *Case where* $x^*(\alpha) = 1$. We know that $y^*(K_1^a) + y^*(K_2^a) + \cdots + y^*(K_l^a) = c_\alpha$. We shall take $\xi_a + \xi_b = c_\alpha$.

This linear system is of the transportation type. If its right-hand side is integer, an integer solution to the linear system may be found easily. It remains to determine ξ_a and ξ_b in order that the solutions x and y found, satisfy the complementary slackness conditions.

Let us construct from $B = (V_1 \cup V_2, A)$ a new bipartite graph $B' = (V_1' \cup V_2', A')$.

$$V_2' = V_2 \cup \{\beta\}, \qquad V_1' = V_1, \qquad A' = A \cup ((\beta, a), (\beta, b)).$$

We define $c_\beta = c_\alpha$.

Solve the maximum weighted stable set in B' and let x' and y' be the integer optimal pair of dual solutions obtained. There exists in B' a stable set of maximum weight containing β, $S^B_{a\bar{b}} \cup \{\beta\}$, and such that $s^{B'}_{\max} = s^B_{\max}$.

The complementary slackness theorem applied to B' allows us to affirm:

- $y'((\beta, a)) + y'((\beta, b)) = c_\alpha$.
- only the edges of the bipartite graph B' with one of their extremities always in a maximum weighted stable set, have a possible positive component.

We shall define: $\xi_a = y'((\beta, a))$, $\xi_b = y'((\beta, b))$.

It is easily proved that the vector x previously described and the vector y, obtained by solving linear system (1) and restricting y' on the edges of B, fulfills the complementary slackness conditions.

(ii) *Case where* $x^*(\alpha) = 0$. This case is identical to the preceding case in everything, except for the quantity

$$y^*(K_1^\alpha)+y^*(K_2^\alpha)+\cdots+y^*(K_l^\alpha),$$

which can be greater than c_α and no longer equal. We shall substitute in this case ξ_a and ξ_b by any integer solution of the following linear system of inequalities:

$$\xi_a \geqslant y'((\beta, a)), \qquad \xi_b \geqslant y'((\beta, b)),$$
$$\xi_a + \xi_b = y^*(K_1^\alpha)+y^*(K_2^\alpha)+\cdots+y^*(K_l^\alpha).$$

In both cases, we were able to obtain a pair of dual integer solutions x and y satisfying the complementary slackness conditions.

Remark 8. In the evaluation of the running time of the proposed algorithm we consider that a sequence of operations Φ_1, Φ_2 and Φ_3 is known, which constructs G starting with a vertex. The most costly procedure contained in the algorithm is the determination of a maximum weighted stable set in a bipartite graph (it is bounded by the cube of the number of vertices of the bipartite graph). The total sum of the vertices from the used bipartite graphs in that construction is less than $|V|$. Hence the complexity of the algorithm is $O(|V|^3)$. This remark uses the fact that the weight of the created vertices during the algorithm are necessarily of the same order as the original weights.

5.2. Maximum weighted clique—minimum covering by maximal cliques

Notations used here are very similar to those in the preceding paragraph.

5.2.1. False twins a and b

Identical with 5.2.3.

5.2.2. True twins a and b

We define $c_\alpha = c_a + c_b$.

(i) Case where $x^*(\alpha)=1$, then $x(a)=1$ and $x(b)=1$.

(ii) Case where $x^*(\alpha)=0$, then $x(a)=0$ and $x(b)=0$.

By a similar procedure to 5.1.1 we obtain in both cases an integer vector y which satisfies with x the complementary slackness theorem.

5.2.3. Extension by a bipartite graph

It is easy to see that the maximum weighted clique is contained either in G^* or in B. We shall restrict ourselves to the case where the maximum weighted clique is in G^*.

We define

$$c_\alpha = \max(c_a, c_b) = c_a.$$

We find the maximum weighted clique in G^*, and in B. We take for the primal solution x the characteristic vector on $G = (V, E)$ of the clique which has the heavier weight. In order to find a dual solution y defined on G, we shall proceed as in 5.1.3 by starting from a dual solution for the maximum weighted clique problem on graph B'. (B' is defined as before, the vertex β of B' supports the weight of the maximum clique defined by x and reduced by c_a.) The dual solution y' on B' has strictly positive components on stable sets of B which are induced by stable sets of B'. The latter ones can be partitioned, due to the fact that they contain, or do not contain, the vertices a and b, in three families, denoted by

$$\mathscr{S}^B_{ab}, \qquad \mathscr{S}^B_{a\bar{b}}, \qquad \mathscr{S}^B_{\bar{a}\bar{b}}.$$

The dual solutions y^* on graph G^* have strictly positive components on stable sets which can be partitioned into two families denoted by

$$\mathscr{S}^{G^*}_\alpha \quad \text{and} \quad \mathscr{S}^{G^*}_{\bar{\alpha}}.$$

A value is assigned to each of the stable sets of these families. This value is initially the (integer) component of vector y' or vector y^* corresponding to the stable set.

We shall now juxtapose two stable sets: one from a family defined in B, the other from a family defined in G^*, in such a way as to obtain a stable set defined in G. Each one of the stable sets thus obtained will correspond to a component of the vector y which is sought. This component is defined by the minimum of the values associated to the two stable sets forming it. We now deduct the value of this minimum from the quantities corresponding to the two stable sets. The stable sets associated with a zero value are then removed.

We continue to define y in the same way, for the remainder of the families. It is, however, necessary to proceed in a fixed order.

We first juxtapose the family $\mathscr{S}^B_{ab}$ with the family $\mathscr{S}^{G^*}_\alpha$ (α is replaced by a and b), then the family $\mathscr{S}^B_{a\bar{b}}$ with $\mathscr{S}^{G^*}_\alpha$ (α is now replaced by a). At this stage, the dual constraints corresponding to vertices a and b are satisfied. We can then juxtapose $\mathscr{S}^{G^*}_\alpha$ (α is removed) with $\mathscr{S}^B_{\bar{a}\bar{b}}$ and finally juxtapose the rest of the latter with $S^{G^*}_{\bar{\alpha}}$.

This procedure allows us to find integer dual solution y which satisfies with x the complementary slackness theorem.

Remark 9. Section 5.2.3 supposes an algorithm is known for the weighted coloring problem for a bipartite graph (dual problem of maximum weighted clique). Such an algorithm, apparently never described, is not difficult to imagine in $O(n^3)$. Using arguments like those of Remark 8, the complexity of the algorithm proposed in Section 5.2 is $O(|V|^3)$.

The very short and technical final Section 5.2 obscures the fact that the problems 'maximum cardinality clique' and 'minimum coloring' have an almost 'greedy' solution. Yet, here again, we find that a parity graph is very much like a bipartite graph: the maximum cardinality clique problem is easier than the maximum cardinality independent set problem.

Acknowledgement

The authors wish to acknowledge Dr. Lex Schrijver and an unknown referee for several very helpful suggestions.

References

[1] C. Berge, Graphes et Hypergraphes (Dunod, Paris, 1970); Graphs and Hypergraphs (North-Holland, Amsterdam, 1973) (English translation).
[2] M. Boulala and J.P. Uhry, Polytopes des indépendants d'un graphe série-parallèle, Discrete Math. 27 (1979) 225–243.
[3] D.G. Corneil, H. Lerchs and L. Stewart, Cographs: a new class of perfect graphs, Discrete Appl. Math., to appear.
[4] L.R. Ford and D.R. Fulkerson, Flows in Networks (Princeton University Press, Princeton, 1962).
[5] T. Gallai, Graphen mit triangulierbaren ungeraden Vielecken, Magyar Tud. Akad. Mat. Kutato Int. Közl. A 7 (1962) 3–36.
[6] M.C. Golumbic, Algorithmic Graph Theory and Perfect Graphs (Academic Press, New York, 1980).
[7] M. Grötschel, L. Lovász and A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, Combinatorica 1 (1981) 169–197.
[8] Wen Lian Hsu and G.L. Nemhauser, Algorithms for minimum coverings by cliques and maximum cliques in claw-free perfect graphs, Cornell University Ithaca, NY, Techn. Rept. no. 418, (1979).
[9] Wen Lian Hsu, How to color claw-free perfect graphs, Cornell University Ithaca, NY, Techn. Rept. no. 435, 1979.
[10] L. Lovász, Normal hypergraphs and the perfect graph conjecture, Discrete Math. 2 (1972) 253–267.
[11] H. Meyniel, On the perfect graph conjecture, Discrete Math. 16 (1976) 339–342.
[12] G.J. Minty, On maximal independent sets of vertices in claw-free graphs, J. Combin. Theory Ser. B 28 (1980) 284–304.
[13] H. Sachs, On the Berge conjecture concerning perfect graphs, in: Combinatorial Structures and their Applications (Gordon and Breach, New York, 1970) pp. 377–384.

[14] L. Surányi, The covering of graphs by cliques, Studia Sci. Math. Hungar. 3 (1968) 345–349.
[15] A.C. Tucker, The strong perfect graph conjecture for planar graphs, Canad. J. Math. 25 (1973) 103–114.
[16] A.C. Tucker, Critical perfect graphs and perfect 3-chromatic graphs, J. Combin. Theory Ser. B 23 (1977) 143–149.

Annals of Discrete Mathematics 16 (1982) 27–55

THE TRAVELLING SALESMAN POLYTOPE AND {0, 2}-MATCHINGS

Gérard CORNUÉJOLS*

Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213, USA

William PULLEYBLANK**

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

A $\{0, 2\}$-matching is an assignment of the integers $0, 2$ to the edges of a graph G such that for every node the sum of the integers on the incident edges is at most two. A tour is the (0–1)-incidence vector of a hamilton cycle. We study the polytope $P(G)$, defined to be the convex hull of the $\{0, 2\}$-matchings and tours of G. When G has an odd number of nodes, the travelling salesman polytope, the convex hull of the tours, is a face of $P(G)$. We obtain the following results:

(i) We completely characterize those facets of $P(G)$ which can be induced by an inequality with (0–1)-coefficients.

(ii) We prove necessary properties for any other facet inducing inequality, and exhibit a class of such inequalities with the property that for any pair of consecutive positive integers, there exists an inequality in our class whose coefficients include these integers.

(iii) We relate the facets of $P(G)$ to the facets of the travelling salesman polytope. In particular, we show that for any facet F of the travelling salesman polytope, there is a unique facet of $P(G)$ whose intersection with the travelling salesman polytope is exactly F.

1. Introduction

Let $G = (V, E)$ be a finite undirected graph and let $c = (c_j : j \in E)$ be a vector of real edge costs. The infamous travelling salesman problem is to find a hamilton cycle of G, the sum of whose edge costs is minimum. (If G has no hamilton cycle, this fact should be discovered.) A major obstacle to be overcome in this process is the verification of a proposed optimum tour. Indeed, even if one has discovered the optimal tour, but is forced to convince a nonbeliever of the tour's optimality, it is generally necessary to perform a

* Research supported, in part, by NSF grant ENG 79-02526 to Carnegie Mellon University.

** Supported by the University of Bonn (Sonderforschungsbereich 21 (DFG), Institut für Operations Research). On leave of absence from Department of Computer Science, The University of Calgary, Canada. Research supported, in part, by National Science and Engineering Research Council of Canada.

quantity of work effectively as large as that performed in finding the optimum tour in the first place.

This need for a good optimality condition prompted the development of the area of polyhedral combinatorial optimization. This approach was pioneered by Jack Edmonds in solving matching problems [4], matroid optimization problems [5, 6] and, as a special case, branching problems [7]. The idea is to represent the feasible solutions to a discrete optimization problem by their incidence vectors and consider the convex hull of these vectors viewed as points of $\mathbb{R}^n$. If a linear system sufficient to define such a polyhedron can be discovered, then linear programming duality theory provides a general min–max optimality criterion.

So far, the results obtained using this approach have not been as successful for the travelling salesman problem as for these other problems. At present no complete characterization of a linear system sufficient to define the convex hull of the set of incidence vectors of the hamilton cycles—the so called travelling salesman polytope—is known. An extensive study of this polytope was carried out by Grötschel [8] as a part of his doctoral dissertation. This continued earlier work of Chvátal [1] who introduced the class of 'comb inequalities', which were generalized ([8, 11]) to provide what is at present the largest known class of essential inequalities for the travelling salesman polytope. Other classes of essential inequalities for the travelling salesman polytope have been obtained by Grötschel [9] and Maurras [14]. However, even though the incompleteness of these linear systems is unsatisfactory from a theoretical point of view, these partial systems have provided the basis for successful cutting plane approaches to 'real world' problems (see [2, 10, 15]).

There is a problem that arises when dealing with the travelling salesman polytope that does not arise when dealing with the polyhedra of matroids and matchings: The travelling salesman polytope is not full dimensional. This means that there does not exist a unique (up to a positive multiple) minimal defining linear system as there does for these other polytopes. In fact an inequality can always be replaced by another obtained by multiplying by a positive constant and adding any linear combination of the equations which define the affine space containing the polytope. Generally, full dimensional polytopes seem more pleasant to handle, so what is often done when studying the travelling salesman polytope is to consider, in fact, the so called *monotone travelling salesman polytope*: the convex hull of the incidence vectors of the hamilton cycles and all subsets of hamilton cycles of a graph. Then the travelling salesman polytope is a face of this larger polytope, the face obtained by requiring $\Sigma\,(x_j : j \in E) = |V|$.

We study here a different full dimensional extension of the travelling salesman polytope. If the number of nodes of G is odd, then, again, the travelling salesman polytope is a proper face of this larger polytope. In the next section we introduce this polytope and completely characterize all the essential inequalities of a defining linear system which can be scaled so as to have (0–1)-coefficients. In Section 3 we prove several necessary properties of any non 0–1 essential inequalities, and give a class of such inequalities. These inequalities have the property that for G sufficiently large, any desired consecutive pair of positive integers can be obtained as coefficients. Finally, in Section 4, we relate the results of Section 3 to the previously known classes of essential inequalities for the travelling salesman polytope. In particular, we show that there exists a natural injection of the set of facets of the travelling salesman polytope into the set of facets of our polytope.

One point of terminology should be clarified at this point. A *facet* of a polytope P is a maximal nonempty proper face, that is, a face of dimension one less than that of P. A *facet inducing inequality* is any inequality which is satisfied by all members of P, and satisfied with equality by precisely the members of some facet F or P. For general polytopes, if we have a minimal defining linear system, then there will be exactly one facet inducing inequality for each facet of P. If the polytope is of full dimension, then every inequality in the system will be facet inducing. However if the polytope is not of full dimension, then this minimal defining system will also include sufficient linear equalities or inequalities to determine the affine space containing the polytope.

2. Tours and {0, 2}-matchings

For any edge j of $G = (V, E)$ we let $\psi(j)$ denote the two nodes of G incident with j. For any $S \subseteq V$ we let $\delta(S)$ denote the set of edges having exactly one end in S and we let $\gamma(S)$ denote the set of edges having both ends in S. We abbreviate $\delta(\{v\})$ by $\delta(v)$ for $v \in V$. For any $J \subseteq E$ and any vector $x = (x_j : j \in E)$ we let $x(J) = \Sigma\,(x_j : j \in J)$. If K is any graph, we will sometimes use $E(K)$ and $V(K)$ to denote the edge set and node set respectively of K.

Now consider the following linear system:

$$0 \leqslant x_j \leqslant 1 \quad \text{for all } j \in E, \tag{2.1}$$

$$x(\delta(i)) \leqslant 2 \quad \text{for all } i \in V, \tag{2.2}$$

$$x(\gamma(S)) \leqslant |S| - 1 \quad \text{for all } S \subset V,\ |S| \geqslant 3\,. \tag{2.3}$$

We define a *tour* to be the incidence vector of the edges of a hamilton cycle of G. It is easily verified that the *integer* solutions to (2.1)–(2.3) are the tours of

G and all incidence vectors of collections of node-disjoint paths. Moreover, if the inequalities in (2.2) are replaced by equations, then the integer solutions are precisely the tours. (This latter system is one of the earliest integer programming formulations of the travelling salesman problem [3].) The constraints (2.2) are called *degree constraints*; the constraints (2.3) are called *subtour elimination constraints.*

Now suppose we remove the upper bound from (2.1). That is, we replace it by

$$0 \leq x_j \quad \text{for all } j \in E. \tag{2.4}$$

The set of 0–1 valued solutions obviously remains unchanged but the set of integer solutions is greatly enlarged. A *1-matching* of G is a set of edges meeting each node at most once. We say that it is *perfect* if it meets each node exactly once. Let M be a 1-matching of G and let $x = (x_j : j \in E)$ be defined by

$$x_j = \begin{cases} 0 & \text{if } j \in E - M, \\ 2 & \text{if } j \in M. \end{cases}$$

We call such a vector a *$\{0, 2\}$-matching* of G and we let $P(G)$ denote the convex hull of the tours and $\{0, 2\}$-matchings of G.

If x is a $\{0, 2\}$-matching of G, then x satisfies (2.2), (2.4) and (2.3) for all those $S \subseteq V$ such that $|S|$ is odd. Of course, every tour will also satisfy these inequalities, and in fact $P(G)$ is the convex hull of the *integer* solutions to this system. This is because every such integer solution other than a tour or a $\{0, 2\}$-matching can be written as $0.5x_1 + 0.5x_2$ where x_1 and x_2 are $\{0, 2\}$-matchings.

In the case that $|V|$ is even, any tour can also be expressed as $0.5x_1 + 0.5x_2$, choosing x_1 and x_2 as the two complementary perfect $\{0, 2\}$-matchings contained in the edges of the tour. Thus when $|V|$ is even, the vertices of $P(G)$ are just the $\{0, 2\}$-matchings. However, when $|V|$ is odd, the situation is quite different. Let TSP(G) denote the travelling salesman polytope of G, i.e., the convex hull of the set of tours of G. Then TSP(G) is the face of $P(G)$ obtained by taking the intersection of $P(G)$ with the affine space defined by

$$x(\delta(i)) = 2 \quad \text{for all } i \in V. \tag{2.5}$$

This is because a graph with an odd number of nodes cannot have a perfect 1-matching and, therefore, if x is a $\{0, 2\}$-matching of G, then there must exist at least one node v for which $x(\delta(v)) < 2$. Conversely, every tour of G satisfies (2.5). Therefore, our objective in this section is to determine several classes of facets of $P(G)$ which we will then relate to TSP(G).

For any $S \subseteq V$ we let $G[S]$ denote the node induced subgraph of G induced by S. We say that $S \subseteq V$ is *hypomatchable* (or 1-critical) if for every $v \in S$, the graph $G[S-\{v\}]$ has a perfect 1-matching. Necessarily, this implies that $|S|$ is odd. Let $Q = \{S \subseteq V\colon S$ is hypomatchable and $|S| \geqslant 3\}$ and let $M(G)$ be the convex hull of the incidence vectors of the 1-matchings of G. Edmonds [4] proved the following theorem.

Theorem 2.1.

$$M(G) = \{x \in \mathbb{R}^E\colon \begin{cases} x_j \geqslant 0 \quad \textit{for all } j \in E, & (2.6) \\ x(\delta(i)) \leqslant 1 \quad \textit{for all } i \in V, & (2.7) \\ x(\gamma(S)) \leqslant (|S|-1)/2 \quad \textit{for all } S \in Q\}. & (2.8) \end{cases}$$

(In fact, the theorem as stated by Edmonds had Q equal to the set of all odd cardinality subsets of V. However, the restriction to hypomatchable sets is implicit in his algorithm used to prove the theorem.) This system of inequalities is 'almost' minimal. Pulleyblank and Edmonds [16] showed that all the inequalities (2.6) are necessary, all the inequalities (2.7) which do not violate a rather technical condition are necessary and an inequality (2.8) is necessary if and only if $G[S]$ is nonseparable, i.e., contains no cut node.

Since a vector x is a {0, 2}-matching if and only if $x/2$ is the incidence vector of a 1-matching, a linear system sufficient to define the convex hull of the set of 2-matchings of G can be obtained by simply doubling the right-hand sides of the linear system (2.6)–(2.8), and trivially, this linear system defines $P(G)$ for $|V|$ even. But when $|V|$ is odd, there is an inequality of the form (2.8) which requires $x(E) \leqslant |V|-1$, and of course, every tour of G violates this inequality.

It is worth noting at this point that a travelling salesman problem on a graph $G = (V, E)$ with $|V|$ even can easily be transformed into an equivalent problem on a graph $G' = (V', E')$ with $|V'|$ odd. Simply duplicate some node v and its incident edges, giving each duplicate edge the same cost as the original. Join v and its duplicate with a simple path having two new nodes and three edges of cost zero. The solution of the travelling salesman problem for G' yields the solution for G.

Since $P(G)$ contains all {0, 2}-matchings of G, $P(G)$ is of full dimension. Therefore for each facet F of $P(G)$ there exists a unique (up to a positive multiple) inequality $ax \leqslant \alpha$ such that $F = \{x \in P(G)\colon ax = \alpha\}$ and every $x \in P(G)$ satisfies $ax \leqslant \alpha$. Moreover, the set of all such inequalities is the minimal defining linear system which we would like to find. Unfortunately, we are unable to explicitly describe this system, but in the following three propositions we define three classes of such facet-inducing inequalities. We will then show that every facet-inducing inequality with 0–1 coefficients belongs to one of these classes.

Proposition 2.2. *For every $j \in E$, $x_j \geqslant 0$ induces a facet of $P(G)$.*

Proof. Let $\mathbf{0}$ denote the zero vector indexed by E and let u^k, for $k \in E$, denote the vector which is zero everywhere but the kth coordinate and $u^k_k = 2$. Then $\{\mathbf{0}\} \cup \{u^k : k \in E - \{j\}\}$ is a set of $|E|$ affinely independent vectors satisfying $x_j = 0$. Since $\{x \in P(G): x_j = 0\}$ is a proper face of $P(G)$, the dimension of this face is $|E| - 1$ and the result follows. □

It is clear that those graphs G which have isolated nodes are uninteresting from a point of view of $P(G)$, since their deletion leaves the polytope unchanged. Henceforth we will always assume that G has no isolated nodes, however, G need not be connected. Of course, if G is not connected, then there exist no tours so the results really reduce to results on $M(G)$.

Proposition 2.3. *For every $v \in V$, $x(\delta(v)) \leqslant 2$ does not induce a facet of $P(G)$ if and only if v has a single neighbor $w \in V$ and $\delta(w) \neq \delta(v)$.*

Proof. If v has a single neighbor u, then any $x \in P(G)$ satisfying $x(\delta(v)) = 2$ also satisfies $x(\delta(u)) = 2$. If there exists $k \in \delta(u) - \delta(v)$, then the unit vector u^k defined in the proof of Proposition 2.2 satisfies $u^k(\delta(u)) = 2$ but $u^k(\delta(v)) = 0$. Therefore $\{x \in P(G): x(\delta(v)) = 2\}$ is not a maximal proper face of $P(G)$ and so $x(\delta(v)) = 2$ does not induce a facet.

Conversely, suppose that v has a single neighbor w but $\delta(w) = \delta(v)$. Let $h \in \delta(v)$. For any $j \in E - \delta(v)$ let $\tilde{u}^j$ be defined by

$$\tilde{u}^j_k = \begin{cases} 0 & \text{if } k \notin \{h, j\}, \\ 2 & \text{if } k \in \{h, j\}. \end{cases}$$

Then $(\tilde{u}^j : j \in E - \delta(v)\} \cup \{u^k : k \in \delta(v)\}$ is a set of $|E|$ affinely independent $\{0, 2\}$-matchings of G, all satisfying $x(\delta(v)) = 2$. Therefore $x(\delta(v)) \leqslant 2$ induces a facet of $P(G)$.

Finally suppose that v has more than one neighbor. Then for any $j \in E - \delta(v)$ there exists a $\{0, 2\}$-matching $\tilde{u}^j$ which is zero everywhere except for the jth component and one component corresponding to a member of $\delta(v)$. Then, as before, $\{\tilde{u}^j : j \in E - \delta(v)\} \cup \{u^k : k \in \delta(v)\}$ is a set of $|E|$ affinely independent $\{0, 2\}$-matchings of G satisfying $x(\delta(v)) = 2$, so $x(\delta(v)) \leqslant 2$ induces a facet of $P(G)$. □

In fact, Propositions 2.2, 2.3 and 2.6 follow immediately from the facet characterizations [16] of the matching polytope $M(G)$. For suppose that $ax \leqslant \alpha$ is a facet inducing inequality for $M(G)$ and that $ax \leqslant 2\alpha$ is a valid inequality for $P(G)$. Then there is a set M of $|E|$ affinely independent incidence vectors x of 1-matchings all satisfying $ax = \alpha$. The set $\bar{M} \equiv \{2 \cdot x : x \in M\}$ is then a set of $|E|$ affinely independent {0, 2}-matchings all satisfying $ax = 2\alpha$, which establishes that the inequality is facet inducing for $P(G)$.

If $|V|$ is odd, then, of course, no perfect 1-matching or {0, 2}-matching of G can exist. Thus we define an *np* ('near perfect') 1-*matching* to be a 1-matching incident with every node of V but one.

Similarly, an *np*-{0, 2}-*matching* is a {0, 2}-matching x of G satisfying $x(E) = |V| - 1$. In other words, only one node is unsaturated. Then a graph G is *hypomatchable* if and only if for every $v \in V$, there exists an np-{0, 2}-matching (or a np-1-matching) which leaves v unsaturated. In [16] the following theorem was proved.

Theorem 2.4. *If G is a nonseparable hypomatchable graph, then there exist $|E|$ np-1-matchings of G, whose incidence vectors are affinely independent.*

This result was proved constructively, via an algorithm which actually constructed the np-1-matchings. Using this result, it was then shown that for $S \subseteq V$ such that $|S| \geqslant 3$, $G[S]$ hypomatchable and nonseparable, the inequality $x(\gamma(S)) \leqslant (|S| - 1)/2$ induces a facet of $M(G)$. A shorter, nonconstructive proof of this result has been obtained by Lovász, which we describe here.

Lemma 2.5. *For every $S \subseteq V$ such that $|S| \geqslant 3$ and $G[S]$ is hypomatchable and nonseparable, $x(\gamma(S)) \leqslant (|S| - 1)/2$ induces a facet of $M(G)$.*

Proof (Lovász). Let X be the set of incidence vectors x of 1-matchings of G which satisfy $x(\gamma(S)) = (|S| - 1)/2$. Since the inequality $x(\gamma(S)) \leqslant (|S| - 1)/2$ is easily seen to be satisfied by all members of $M(G)$, all we need show is that the affine rank of X is equal to $|E|$, or in other words, there is a unique (up to a positive multiple) nonzero vector $a = (a_j : j \in E)$ and scalar α such that $ax = \alpha$ for every $x \in X$. To do this, we will show that any such a must satisfy $a_j = k$ for some constant k, for all $j \in \gamma(S)$ and $a_j = 0$ for all $j \in E - \gamma(S)$. For then, if we 'scale' a by dividing every component by k, we see that this inequality must be a scalar multiple of the inequality $x(\gamma(S)) \leqslant (|S| - 1)/2$.

So suppose there exist $i \in S$ such that a_j takes on different values for edges j in $\delta(i) \cap \gamma(S)$. Let the graph G' be obtained from $G[S]$ by 'splitting' i into two nodes i' and i'' such that all the edges j of $\delta(i) \cap \gamma(S)$ for which a_j takes on the minimum value are incident with i' and all the others are adjacent with i''. Since $G[S]$ was nonseparable, G' is connected and in addition must have a perfect 1-matching. For if not, by Tutte's theorem [17] (viz (2.8)), there would exist a nonempty subset Y of the node of G' such that deleting Y creates at least $|Y|+2$ odd cardinality components.

Let Y' be the subset of S consisting of all nodes of $Y \cap S$ together with i if at least one of i', i'' belongs to Y. Then by checking cases one can verify that $G[S-Y']$ has at least $|Y'|$ odd cardinality components. Therefore for any $v \in Y'$, $G[S-\{v'\}]$ does not have a perfect 1-matching, contrary to $G[S]$ being hypomatchable.

So let x^* be the incidence vector of a perfect matching of G', and let $j' \in \delta(i')$ and $j'' \in \delta(i'')$ be such that $x^*_{j'} = x^*_{j''} = 1$. Let x^1 and x^2 be obtained from x^* setting the j' and j'' components respectively to zero. Then x^1, $x^2 \in X$ but $ax^1 > ax^2$, a contradiction to $ax = \alpha$ for all $x \in X$. Therefore, the value of a_j is constant for all $j \in \gamma(S)$. Moreover, it is easily verified that for every $j \in E-\gamma(S)$, there exists $x' \in X$ such that $x'_j = 1$. Moreover, the vector x'' obtained from x' by setting the jth component to 0 also belongs to X. Therefore, we must have $a_j = 0$ for $j \in E-\gamma(S)$ and the result follows. □

Proposition 2.6. *For every $S \subsetneq V$ such that $|S| \geq 3$ and $G[S]$ is hypomatchable and nonseparable, $x(\gamma(S)) \leq (|S|-1)$ induces a facet of $P(G)$.*

Proof. This is an immediate corollary of Lemma 2.5. □

The important difference between Lemma 2.5 and Proposition 2.6 is that in the latter we were forced to restrict S to being a *proper* subset of V, because every tour of G violates the inequality $x(E) \leq |V|-1$. For the case $S=V$, we have the following result for $P(G)$.

Proposition 2.7. *Let $G' = (V, E')$ be a spanning subgraph of G which is hypomatchable, nonseparable and nonhamiltonian and such that E' is maximal with this property. Then $x(E') \leq |V|-1$ induces a facet of $P(G)$.*

Proof. Since G' is nonhamiltonian and $|V|$ is odd, every member of $P(G)$ must satisfy $x(E') \leq |V|-1$. All we need show is that there exist $|E|$ affinely independent members of $P(G)$, all of which satisfy $x(E') = |V|-1$. First we note that since G' is nonseparable and hypomatchable, it follows from Proposition 2.4 that there exists a set X of $|E'|$ affinely independent incidence vectors of np-1-matchings of G'. Let $\bar{X}$ be obtained from X by taking each $x \in X$, doubling it and defining the jth component to be zero for all $j \in E-E'$. Then

$\bar{X}$ is a set of $|E'|$ affinely independent np-$\{0, 2\}$-matchings of G. Moreover, $x_j = 0$ for all $j \in E - E'$, for all $x \in \bar{X}$. For each $j \in E - E'$, there exists a hamilton cycle whose edges are contained in $E' \cup \{j\}$, by the maximality of E'. Let t^j be the tour corresponding to such a hamilton cycle. Then $t^j(E') = |V| - 1$, for all $j \in E - E'$ and we let $T = \{t^j : j \in E - E'\}$. It is easily seen that $T \cup \bar{X}$ is a set of $|E|$ affinely independent members of $P(G)$ all satisfying $x(E') = |V| - 1$, since for $j \in E - E'$, t^j is the only member of $T \cup \bar{X}$ for which the jth component is nonzero. □

We note that if G is nonhamiltonian, then the inequality of the previous proposition is simply $x(E) \leqslant |V| - 1$ which is facet inducing for $P(G)$ if G is nonseparable and hypomatchable. However, when G is hamiltonian, then G' must be a proper spanning subgraph of G, which is therefore *not* node induced. In general, the number of these subgraphs is very large.

We make use of one more preliminary result. For any $X \subseteq V$ let $c(X)$ denote the number of components of $G[V - X]$ having an odd number of nodes. Tutte's classical theorem characterizing those graphs having perfect 1-matchings is the following.

Theorem 2.8 (Tutte [17]). *G has a perfect 1-matching if and only if for every $X \subseteq V$, $|X| \geqslant c(X)$.*

A less classical theorem characterizing those graphs which are hypomatchable was proved independently by Pulleyblank and Edmonds [16] and Lovász [12].

Theorem 2.9. *$G = (V, E)$ is hypomatchable if and only if $|V|$ is odd and for every nonempty $X \subseteq V$, $|X| \geqslant c(X)$.*

Of course, the important part of these theorems is the sufficiency of the condition, i.e., the assertion that if G has *no* perfect matching or is *not* hypomatchable, then there exist $X \subseteq V$ such that $|X| < c(X)$. It is not difficult to strengthen this in the following manner.

Corollary 2.10. *Let $G = (V, E)$ be a connected graph which has no perfect matching and is not hypomatchable. Then there exists nonempty $X^* \subseteq V$ which maximizes $c(X) - |X|$ over all nonempty $X \subseteq V$, and such that $G[V - X^*]$ consists only of (at least $|X^*| + 1$) hypomatchable components.*

Proof. We prove by induction on $|V|$. If $|V| = 1$, then G is hypomatchable; if $|V| = 2$ or 3, then the assertion is easily checked. Suppose G has k nodes and the result is true for all smaller graphs. By Theorems 2.8 and 2.9 there exists nonempty $X \subseteq V$ such that $c(X) - |X| > 0$, let X^* be chosen such that $c(X^*) - |X^*|$ is maximum and, subject to this, the number of nonhypomatchable components of $G[V - X^*]$ is minimum. If there are no such components, then we are done, so suppose that S is the nodeset of a nonhypomatchable component. If $|S|$ is odd, then by induction there is $\emptyset \neq X_S \subseteq S$ such that $G[S - X_S]$ consists of at least $|X_S| + 1$ hypomatchable components. Then

$$c(X^* \cup X_S) - |X^* \cup X_S| \geq c(X^*) - |X^*|,$$

but $G[V - (X^* \cup X_S)]$ contains fewer nonhypomatchable components than does $G[V - X^*]$, a contradiction. If $|S|$ is even, then let v be any node of S, which is not a cutnode of $G[S]$ and let $X' \equiv X \cup \{v\}$. Then

$$c(X') - |X'| = c(X^*) - |X^*|$$

and if $G[S - \{v\}]$ is hypomatchable, then we have contradicted the choice of X^*. If not, then as before we use induction to find $X_S \subseteq S - \{v\}$ such that $X' \cup X_S$ contradicts the choice of X. □

Our next theorem provides a characterization of all those facet inducing inequalities of $P(G)$ which can be scaled so as to have 0–1 coefficients. Thus we say that an inequality $ax \leq \alpha$ is a (0–1)-inequality if every $a_j \in \{0, -1, 1\}$. We will also say that such an inequality is a *(0–k)-inequality* if every $a_j \in \{0, -k, k\}$ for some positive real number k. Then, of course, to any (0–k)-inequality there corresponds a (unique) (0–1)-inequality obtained by multiplying by $1/k$.

Note that this definition allows us to consider the inequality $x_j \geq 0$ for $j \in E$ (equivalently, $-x_j \leq 0$) as a (0–1)-inequality. It might be asked whether there exist other facet inducing inequalities $ax \leq \alpha$ for $P(G)$ having $a_j < 0$ for some j. We can answer this in the negative; all others are obtained from nonnegativity constraints by scaling.

Lemma 2.11. *If $ax \leq \alpha$ is a facet inducing inequality for $P(G)$ having $a_j < 0$ for some $j \in J$, then this inequality must be $a_j x_j \leq 0$.*

Proof. Since $P(G)$ is of full dimension, if we let M be the set of $\{0, 2\}$-matchings x of G satisfying $ax = \alpha$ and let T be the set of tours t of G satisfying $at = \alpha$ then the affine rank of $X \equiv M \cup T$ must be $|E|$. Therefore $ax = \alpha$ is the unique hyperplane containing all elements of X. Suppose $a_j < 0$. If there existed $\hat{x} \in M$ for which $\hat{x}_j > 0$, then setting the jth component of $\hat{x}$ to 0 would yield another $\{0, 2\}$-matching of G violating $ax \leq \alpha$. If there existed

$t \in T$ with $t_j > 0$, then setting the jth component to zero gives the incidence vector x of a hamilton path of G, for which $ax > \alpha$. But x is the average of two $\{0, 2\}$-matchings of G, at least one of which must violate $ax \leqslant \alpha$. Thus we must have $x_j = 0$ for all $x \in X$ so $ax \leqslant \alpha$ is a positive multiple of the nonnegativity constraint $-x_j \leqslant 0$. □

For any $J \subseteq E$ we let $r(J)$ denote the maximum possible value for $x(J)$ for all $\{0, 2\}$-matchings and tours x of G. This 'rank' function is important for it is the smallest possible value for α if the (0–1)-inequality $x(J) \leqslant \alpha$ is to be valid for $P(G)$. Moreover, if $\alpha > r(J)$, then no member of $P(G)$ can satisfy $x(J) = \alpha$. Thus $r(J)$ is the only possible value for α if $x(J) \leqslant \alpha$ is to induce a facet of $P(G)$.

Finally, let W denote the set of all $v \in V$ such that either v has at least two neighbors, or, if v as a single neighbor w, then $\delta(v) = \delta(w)$.

Theorem 2.12. *The following is the complete set of facet inducing* (0–1)-*inequalities of* $P(G)$:

$$x_j \geqslant 0 \quad \text{for all } j \in E, \tag{2.9}$$

$$x(\delta(i)) \leqslant 2 \quad \text{for all } i \in W, \tag{2.10}$$

$$x(\gamma(S)) \leqslant |S| - 1 \quad \text{for all } S \subsetneq V,\ |S| \geqslant 3,\ G[S] \text{ nonseparable, hypomatchable,} \tag{2.11}$$

$$x(E') \leqslant |V| - 1 \quad \text{for all edge maximal spanning subgraphs } G' = (V, E') \text{ of } G \text{ which are hypomatchable, nonhamiltonian and nonseparable.} \tag{2.12}$$

Proof. We saw in Propositions 2.2, 2.3, 2.6 and 2.7 that all these inequalities do induce facets of $P(G)$. Now we show that every facet inducing (0–1)-inequality is of one of the above types. Let $ax \leqslant \alpha$ be facet inducing. By Lemma 2.11, if $ax \leqslant \alpha$ is not of the form (2.9), we must have $a \geqslant 0$, so let $E' \equiv \{j \in E\colon a_j = 1\}$. Then the inequality $ax \leqslant \alpha$ must be $x(E') \leqslant r(E')$. Edmonds observed, in the context of matroid polyhedra, that if such an inequality is facet inducing, then two properties must hold: First, E' must be *closed*, i.e., for every $j \in E - E'$, we must have $r(E' \cup \{j\}) > r(E')$. Otherwise $x(E' \cup \{j\}) \leqslant r(E' \cup \{j\}) = r(E')$ would be a stronger valid inequality than $ax \leqslant \alpha$, contradicting the necessity of a facet-inducing inequality. Second, E' must be *nonseparable*, i.e., there cannot exist nonempty S, $T \subseteq E'$ such that $S \cup T = E'$ and $r(S) + r(T) = r(E')$. For in this case, the inequality $x(E') \leqslant r(E')$ is implied by the sum of the inequalities $x(S) \leqslant r(S)$ and $x(T) \leqslant r(T)$ which means that it can be replaced with these two different inequalities, again contradicting the necessity. So we know that E' is closed and nonseparable.

Let V' be the set of nodes incident with edges in E'. If the graph $G' = (V', E')$ has a perfect $\{0, 2\}$-matching or contains a hamilton cycle of G, then $r(E') = |V'|$ and so $ax \leqslant \alpha$ is implied by one half of the sum of the degree constraints for the nodes in V'. Since the degree constraints are valid inequalities the facet $ax \leqslant \alpha$ can be nécessary only if it is identical to a degree constraint (2.10). Furthermore the assumption that G' has a perfect $\{0, 2\}$-matching means that $V' = \{u, v\}$ and $\delta(u) = \delta(v) = E'$.

Now suppose that G' is hypomatchable. Then $r(E') = |V'| - 1$. If $V' \subset V$, then since E' is closed we must have $E' = \gamma(V')$. If G' were separable, then E' would be separable, so we must have $G' = G[V']$ is hypomatchable, nonseparable, so that $ax \leqslant \alpha$ is an inequality of the form (2.11). If $V' = V$, then G' is a hypomatchable non-hamiltonian spanning subgraph of G which must also be nonseparable and edge maximal with these properties, since E' is closed and nonseparable. Thus $ax \leqslant \alpha$ is a constraint of the form (2.12).

Finally, suppose that G' has no perfect matching and is not hypomatchable. By Corollary 2.10 there exist nonempty $X \subseteq V'$ such that $G'[V' - X]$ consists of at least $|X| + 1$ hypomatchable components, and $c'(X) - |X|$ is maximized, where $c'(X)$ denotes the number of odd components of $G'[V' - X]$. If we sum the degree constraints (2.10) for the nodes of X and the constraints (2.11) for the node sets of the components of $G'[V' - X]$ (or the nonseparable blocks of these components if they contain cutnodes), then we obtain a valid inequality $x(\bar{E}) \leqslant |V'| - (c'(X) - |X|)$, where $\bar{E} \supseteq E'$. (If some of these components are single nodes, the constraint (2.11) is trivial and can be dropped.) Now if $|X| = 1$ and every component of $G'[V' - X]$ consists of a single node, then $ax \leqslant \alpha$ must be a degree constraint (2.10). Otherwise the inequality $x(\bar{E}) \leqslant |V'| - (c'(X) - |X|)$ has been derived as the sum of at least two different valid inequalities. We will show that $r(E') = |V'| - (c'(X) - |X|)$ which will contradict $x(E') \leqslant r(E')$ being facet, since we can obtain it (or a stronger inequality) from other inequalities.

Clearly $r(E') \leqslant |V'| - (c'(X) - |X|)$; all we need do is find some $x^* \in P(G)$ giving equality. Construct a bipartite graph $\tilde{G}$ from G' having one node $v(x)$ for each $x \in X$, one node $v(K)$ for each component K of $G'[V' - X]$ and an edge joining $v(x)$ and $v(K)$ if and only if x was adjacent (in G') to some node of K. If there is no 1-matching which covers all nodes $v(x)$ for $x \in X$, then by Hall's theorem, there is a set $\tilde{X} \subseteq X$ such that fewer than $|\tilde{X}|$ nodes $v(K)$ for components K of $G'[V' - X]$ are adjacent to nodes $v(x)$ for $x \in \tilde{X}$. But then $c'(X - \tilde{X}) - |X - \tilde{X}| > c'(X) - |X|$ a contradiction. So we can construct x^* by letting $x_j^* = 2$ for each edge corresponding to an edge of a maximum 1-matching of $\tilde{G}$, by letting x_j^* be defined equal to an appropriate np-$\{0, 2\}$-matching for each component of $G'[V'-X]$ and $x^*_j = 0$ otherwise. Then $x^* \in P(G)$ and $x^*(E') = 2|X| + \Sigma\{|S|-1$: S is the nodeset of a component of $G'[V'-X]\} = |V'| - (c'(X) - |X|)$. □

3. General facets of $P(G)$

We discuss three topics in this section. First we establish some necessary conditions which must be satisfied by any facet inducing non-(0–1)-inequality of $P(G)$. Second, we describe such a class of non-(0–1)-inequalities having the following property: For any pair $(s, s+1)$ of consecutive positive integers, there exists a graph G and a facet inducing inequality whose coefficients include s and $s+1$. Third we describe a lifting procedure which allows us to obtain facets of $P(G)$ from facets of $P(G')$ for a subgraph G' of G.

As we remarked in Section 2, if G has an even number of nodes, then tours are the midpoints of lines joining pairs of {0, 2}-matchings and so by Theorem 2.1 $P(G)$ has only 0–1 facets. Thus the only graphs we will consider in the remainder of this paper are those having an odd number of nodes.

Let $ax \leqslant \alpha$ be an integer inequality. We say that this inequality is non-(0–1) if it is impossible to scale the coefficients so that $a_j \in \{0, \pm 1\}$ for all $j \in E$. In other words, there exist two nonzero coefficients with different magnitudes. If t is a tour of $G = (V, E)$, we let $E(t) \equiv \{j \in E\colon t_j = 1\}$.

Theorem 3.1. *Let G have an odd number of nodes and let $ax \leqslant \alpha$ be a non-(0–1) facet inducing inequality for $P(G)$. Let $E^+ \equiv \{j \in E\colon \alpha_j \neq 0\}$ and let G^+ be the subgraph of G induced by the edges in E^+. Let M be the set of {0, 2}-matchings x satisfying $ax = \alpha$ and let T be the set of tours t satisfying $at = \alpha$. Then*

$$a_j > 0 \quad \textit{for all } j \in E^+ \textit{ and } \alpha > 0\,, \tag{3.1}$$

every $x \in M$ is a np-matching of G, (3.2)

for every node v of G, M contains a np-matching deficient at v, (3.3)

any basis of $T \cup M$ contains at least one tour t for which $E(t) \subseteq E^+$; there exists a basis B of $T \cup M$ such that every $t \in B$ satisfies $|E(t) - E^+| \leqslant 1$. (3.4)

Note that (3.4) imples that G^+ is a spanning hamiltonian subgraph of G, which of course implies that G^+ is hypomatchable. Condition (3.3) adds that M contains a np-matching of G deficient at every node of G. Moreover, there exists a basis of $M \cup T$, thus, a subset of the points sufficient to uniquely define the facet, consisting solely of np-matchings and tours which are either contained in G^+ or else induce hamilton paths in G^+.

Proof. Since $P(G)$ is of full dimension, the affine rank of $X \equiv M \cup T$ must be $|E|$, so $ax = \alpha$ is the unique hyperplane containing all elements of X. Since $ax \leqslant \alpha$ is a non-(0–1)-inequality, Lemma 2.11 yields $a \geqslant 0$, which in turn implies $\alpha > 0$, so (3.1) is immediate. Now we show that G^+ must be connected. If not, let K be a component of G^+ and let a' and a'' be defined by

$$a'_j \equiv \begin{cases} a_j & \text{if } j \in E(K), \\ 0 & \text{otherwise}; \end{cases}$$

$$a''_j \equiv \begin{cases} 0 & \text{if } j \in E(K), \\ a_j & \text{otherwise}. \end{cases}$$

If there exists x^1, $x^2 \in X$ such that $a'x^1 > a'x^2$, then we can find such x^1, $x^2 \in M$. For let $t \in T$. Since we assume G^+ is not connected, $E(t) \cap E^+$ will consist of some number of disjoint paths, and consequently can be expressed as the average of two $\{0, 2\}$-matchings x^1, x^2 both of which must be in M. But then we must have $a'x^1 \leqslant a't \leqslant a'x^2$ and so one of x^1, x^2 would serve as a substitute for t. Since, therefore, x^1, $x^2 \in M$ we can define x^* equal to x^1 on $E(K)$ and equal to x^2 on the rest of G and then $ax^* > \alpha$, a contradiction. Therefore

$$G^+ \text{ is connected}. \tag{3.5}$$

If every $x \in M$ satisfied $x(\delta(v)) = 2$ for some $v \in V$, then since every $t \in T$ must satisfy $t(\delta(v)) = 2$, we must have $ax \leqslant \alpha$ being a degree constraint (2.10). Since we have assumed that $ax \leqslant \alpha$ does not induce a 0–1 facet, we must have, therefore,

$$\text{for each } v \in V \text{ there exist } x \in M \text{ such that } x(\delta(v)) = 0. \tag{3.6}$$

Now we show that

$$\text{every } x \in M \text{ satisfies } x(E^+) = |V(G^+)| - 1 \tag{3.7}$$

which will mean that every $x \in M$ induces a np-matching of G^+.

Suppose that some $x \in M$ satisfies $x(\delta(u) \cap E^+) = x(\delta(v) \cap E^+) = 0$ for some u, $v \in V(G^+)$. Assume that x, u and v are chosen so that the distance in G^+ from u to v is minimum. (This is well-defined in view of (3.5).) If u and v were adjacent in G^+, then by defining $x_j \equiv 2$ for an edge j of E^+ joining u and v, we would violate the constraint $ax \leqslant \alpha$. Therefore there exists a node w different from u and v on a shortest path in G^+ from u to v and by (3.6) we can find $\hat{x} \in M$ satisfying $\hat{x}(\delta(w)) = 0$. There exists a path in G^+, starting at w, whose edges alternately have the value two in x and $\hat{x}$. If we replace the values of x_j with those of $\hat{x}_j$ for the edges in the path, then we will obtain a {0, 2}-matching $x^* \in M$ deficient at w and (at least) one of u, v. But this contradicts the minimal distance property of x, u, v. Thus $|V(G^+)|$ is odd and (3.7) is established.

If $T = \emptyset$, or if there exists a basis B of $M \cup T$ which is contained in M, then $ax \leqslant \alpha$ must be the constraint $x(E^+) \leqslant |V(G)| - 1$ since, by (3.7), every element in the basis satisfies it as an equality. This is a 0–1 facet and thus it must be of the type (2.11) or (2.12) by Theorem 2.12. Since we assume $ax \leqslant \alpha$ is not a (0–1)-constraint, therefore, every basis of $M \cup T$ contains at least one tour.

(We remark that to this point this proof parallels a proof of Lovász [13], who gave a nonalgorithmic proof of the sufficiency of the linear system (2.6)–(2.8) for the 1-matching polytope.)

So we must have $T \neq \emptyset$. We will show that

for any tour $t \in T$, either $E(t) \subseteq E^+$ or else $E(t)$ induces a hamilton path in G^+. (3.8)

For suppose $E(t) \nsubseteq E^+$. Then there is $j \in E(t)$ with $a_j = 0$ and $E(t) - \{j\}$ consists of the edges of an even length path of G since G has an odd number of nodes. This path can be expressed as the average of two complementary np-matchings x^1 and x^2 of G, and $at = 0.5ax^1 + 0.5ax^2 = \alpha$ which implies that $x^1, x^2 \in M$ (since every $x \in P(G)$ satisfies $ax \leqslant \alpha$). If $E(t) \cap E^+$ is not a hamilton path of G^+, then it is easily verified that one of x^1, x^2 will violate (3.7). Thus (3.8) is established.

Now in order to complete the proof, we must show that

there exists a basis B^* of $T \cup M$ such that every $t \in B^*$ satisfies $|E(t) - E^+| \leqslant 1$. (3.9)

This will imply that G^+ is a spanning subgraph of G, and hence (3.7) will imply (3.2), which combined with (3.6) will prove (3.3).

Let B be a basis of $M \cup T$ containing a minimum number of tours t for which $|E(t) - E^+| > 1$, and suppose that $\bar{t}$ is such a tour. Then $E(\bar{t})$ induces an (even length) hamilton path π^0 of G^+, and $E(\bar{t}) - E^+$ consists of a single odd length path π^1, which contains an even number of nodes (including the end points u, v which are nodes of G^+) (see Fig. 3.1). Moreover, since $|E(\bar{t}) - E^+| > 1$, we have $|E(\pi^1)| \geqslant 3$. Now for each $w \in V^* \equiv V(\pi^1) - \{u, v\}$, let x^w be the np-matching of G obtained by taking the (unique) perfect matching of the path obtained from $\bar{t}$ by deleting w.

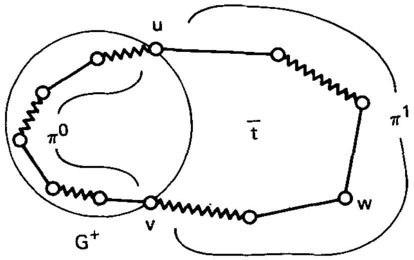


Fig. 3.1.

Let $s \in V(G^+) - \{u, v\}$ and let $\hat{x} \in M$ satisfy $\hat{x}(\delta(s)) = 0$ and $\hat{x}_j = 0$ for all $j \notin E^+$. (By (3.6), there exists $\hat{x} \in M$ satisfying the first property, and we can simply require $\hat{x}_j \equiv 0$ for all $j \in E - E^+$.) Finally, let $\tilde{x}$ be obtained from $\hat{x}$ by giving $\tilde{x}_j$ the value two for the second, fourth, etc. edges of π^1. Then $\tilde{x} \in M$. Now it can be easily verified that

$$|V^*|\bar{t} = \sum (x^w \colon w \in V^*) + \tilde{x} - \hat{x},$$

so $\bar{t}$ is a linear combination of $M^+ \equiv \{x^w \colon w \in V^*\} \cup \{\tilde{x}, \hat{x}\}$. Moreover, $\alpha = a\bar{t} = (1/|V^*|)\Sigma(ax^w \colon w \in V^*) + \alpha - \alpha$ so we must have $\Sigma(ax^w \colon w \in V^*) = |V^*|\alpha$. But since every x^w satisfies $ax^w \leq \alpha$, we must have, therefore, $ax^w = \alpha$ for all $w \in V^*$. Therefore $M' \subseteq M$, and so any basis of $B - \{\bar{t}\} \cup M'$ will be a basis of $M \cup T$ which contradicts our choice of B. Thus (3.9) is established.

Finally, note that if there existed a basis B of $M \cup T$ such that every tour $t \in B$ satisfied $|E(t) - E^+| = 1$, then every $x \in B$ would satisfy $x(E^+) = |V| - 1$, so our constraint would necessarily be the inequality $x(E^+) \leq |V| - 1$ which induces a (0–1)-facet of the form (2.12). Thus (3.4) is established and the proof is complete. □

The consequences of this theorem are quite important. It shows that any facet inducing non-(0–1)-inequality must come from a (spanning) subgraph of G which contains hamilton cycles of G. We now examine such a class of facets.

A simple example of a graph G for which $P(G)$ has a non 0–1 facet is the graph of Fig. 3.2(a). If we let $a = (a_j \colon j \in E)$ be the vector of edge coefficients indicated in the figure, then $ax \leq 14$ is a facet of $P(G)$.

It is easy to verify that the inequality is valid; we show that it is a facet by exhibiting $|E(G)| = 12$ affinely independent members of $P(G)$ satisfying $ax = 14$. These will consist of eleven $\{0, 2\}$-matchings and one tour. In order to obtain the $\{0, 2\}$-matchings, we consider the seven np-$\{0, 2\}$-matchings of the centre heptagon and extend each to a np-$\{0, 2\}$-matching of G by setting $x_j = 2$ for the edge j joining u and v. These are easily seen to be independent and use only the edge j of the graph G^* obtained by contracting the heptagon (see Fig.

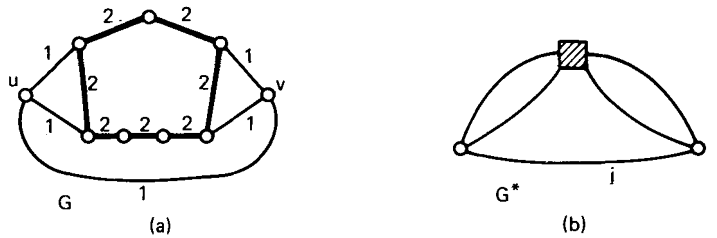


Fig. 3.2. Graph G for which $P(G)$ has a non-(0–1) facet.

3.2(b)). We can give any one of the other four edges of G^* the value two and extend it to a np-$\{0,2\}$-matching of G which is near perfect on the heptagon, and thereby obtain four more. Finally, the unique tour on G is affinely independent from the $\{0,2\}$-matchings and so we are done.

Notice that the idea of the construction was to take a large set of np-$\{0,2\}$-matchings which were also near perfect on a certain induced subgraph, and then complete them with a tour. This provides the basis for a general construction.

Let $G' = (V', E')$ be a subgraph of a hamiltonian graph G. We define

$$\tau(G') \equiv |V'| - \max\{t(E'): t \text{ is a tour of } G\}.$$

We call τ the *segment number* of G'; it equals the smallest number of segments of some hamilton cycle of G which cover all the nodes of G'. For example, if H is the heptagon of Fig. 3.2(a), then $\tau(H) = 2$. If G is nonhamiltonian, then the function τ is not defined.

For any $S \subseteq V$ for a graph $G = \{V, E)$ we let $G \times S$ denote the graph obtained by contracting the subgraph $G[S]$ to a single *pseudonode*. Thus, in Fig. 3.2, $G^* \equiv G \times V(H)$.

Theorem 3.2. *Let $G = (V, E)$ be hamiltonian, let $G' = (V', E')$ be a node induced subgraph of G and suppose that*

$$G' \textit{ is hypomatchable and nonseparable}, \tag{3.10}$$

$$G \times V' \textit{ is hypomatchable and nonseparable}. \tag{3.11}$$

Let $a = (a_j: j \in E)$ be defined by

$$a_j \equiv \begin{cases} \tau(G') & \text{for } j \in E', \\ \tau(G') - 1 & \text{for } j \in E - E', \end{cases}$$

and let

$$\alpha \equiv (\tau(G') - 1)(|V| - 1) + |V'| - 1.$$

Then $ax \leq \alpha$ induces a facet of $P(G)$.

Proof. We first prove the validity of $ax \leqslant \alpha$. If $\bar{x}$ is a $\{0,2\}$-matching of G, then $\bar{x}(E) \leqslant |V| - 1$ and $\bar{x}(E') \leqslant |V'| - 1$ and so $a\bar{x} \leqslant \alpha$. Moreover $a\bar{x} = \alpha$ if and only if $\bar{x}$ is near perfect on both G and G'. If $\bar{x}$ is a tour of G, then $\bar{x}(E) = |V|$ and $\bar{x}(E') \leqslant |V'| - \tau(G')$ so $a\bar{x} = (\tau(G') - 1)\,\bar{x}(E) + \bar{x}(E') \leqslant (\tau(G') - 1)(|V| - 1) + |V'| - 1 = \alpha$, and we have $a\bar{x} = \alpha$ if and only if $\bar{x}(E') = |V'| - \tau(G')$. Since $ax \leqslant \alpha$ is valid for all vertices of $P(G)$, it is valid for $P(G)$.

We show that $ax \leqslant \alpha$ is facet inducing by exhibiting $|E|$ affinely independent members x of $P(G)$ satisfying $ax = \alpha$. By (3.10) and Proposition 2.4, there are $|E'|$ affinely independent np-$\{0,2\}$-matchings of G'. Let $\hat{x}$ be any np-$\{0,2\}$-matching of $G \times V'$ which is deficient at the pseudonode V'. We extend each of our np-$\{0,2\}$-matchings of G' to a np-$\{0,2\}$-matching of G by defining it equal to $\hat{x}$ on $E - E'$. Let X^0 be the set of affinely independent matchings thereby obtained. Then $ax = \alpha$ for all $x \in X^0$.

By (3.11) and Proposition 2.4 there are $|E - E'|$ affinely independent np-$\{0,2\}$-matchings of $G \times V'$. Let $\bar{X}$ be such a set, which contains $\hat{x}$, and let X^1 be obtained from $\bar{X} - \{\hat{x}\}$ by extending each $x \in \bar{X} - \{\hat{x}\}$ to a np-$\{0,2\}$-matching of G. Then X^1 is a set of $|E - E'| - 1$ affinely independent $\{0,2\}$-matchings of G, each $x \in X^1$ satisfies $ax = \alpha$, and it is straightforward to verify that $X^0 \cup X^1$ is affinely independent.

Finally, let t be a tour of G satisfying $t(E') = |V'| - \tau(G')$. Then $at = \alpha$ and since $t(E) = |V|$ but $x(E) = |V| - 1$ for all $x \in X^0 \cup X^1$, we see that $X^0 \cup X^1 \cup \{t\}$ is affinely independent completing the proof. $\square$

Thus, the example of Fig. 3.2 is simply an application of Theorem 3.2, taking G' to be the heptagon. The smallest graph G we know for which $P(G)$ has such a facet inducing non-(0–1)-inequality is the example of Fig. 3.3. We let G' be the triangle, and then $x(E - E') + 2x(E') \leqslant 8$ is facet inducing.

The graph of Fig. 3.4 is a nine node example of a graph for which $P(G)$ has a facet inducing inequality with coefficients 2 and 3. Again G' is the center triangle. Then $\tau(G') = 3$ and $2x(E - E') + 3x(E') \leqslant 18$ is facet inducing.

We obtain a facet inducing inequality of $P(G)$ for a graph G containing any desired consecutive pair $(s, s+1)$ of integers as its nonzero coefficients by a generalization of the construction of Fig. 3.4. Start with an odd polygon P having k nodes, for $k \geqslant s + 1$. Then attach $s + 1$ 'ears'—paths of length three—to adjacent pairs of nodes of P. Finally choose some node v^* which is an

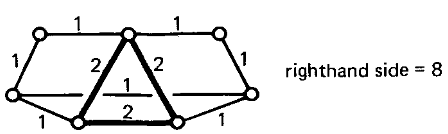


Fig. 3.3. Seven node graph for which $P(G)$ cannot be described by a set of (0–1)-inequalities.

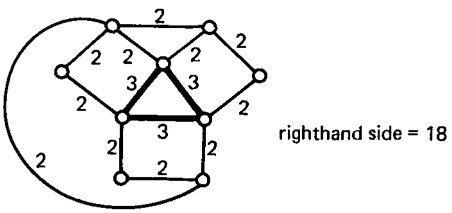


Fig. 3.4.

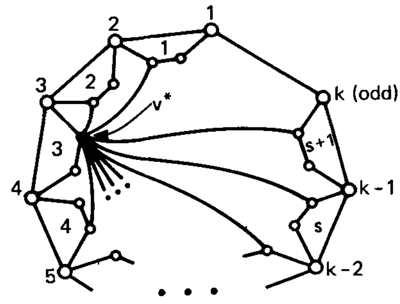


Fig. 3.5.

interior node of some ear. Join v^* to the non-corresponding node of each other ear. Then for this graph G it will follow from Theorem 3.2 that $(s+1)\times x(E(P)) + s\,x(E - E(P)) \leq sk + 2s^2 + s + k - 1$ is facet inducing for $P(G)$ (see Fig. 3.5).

Suppose we have a spanning subgraph $G' = (V, E')$ of G and suppose we know a facet inducing inequality $a'x \leq \alpha'$ for $P(G')$. We will say that a facet inducing inequality $ax \leq \alpha$ for $P(G)$ is obtained by lifting $a'x \leq \alpha'$ if

$$\alpha' = \alpha\,, \qquad \alpha'_j = a_j \quad \text{for all } j \in E'\,.$$

In other words, we do not change the existing coefficients or right-hand side, we simply define those coefficients not previously defined in such a manner that the resulting inequality induces a facet of $P(G)$.

A simple method of obtaining such inequalities is the following *sequential lifting* procedure.

Procedure 3.3 (*sequential lifting*)

Input: $G = (V, E)$, a spanning subgraph $G' = (V, E')$ and a facet inducing inequality $a'x \leq \alpha$ of $P(G')$.

Output: A facet inducing inequality $ax \leq \alpha$ of $P(G)$ obtained by lifting $a'x \leq \alpha$.

Procedure

[0] Initially, define $a_j \equiv a'_j$ for $j \in E'$ and let $S \equiv E'$. S is the set of edges for which a_j has been defined.

[1] For each $j \in E - S$, do the following

[1a] Let u, v, be the ends of j. Let

$a_j \equiv \min\{1/2(\alpha - ax)$: x is a $\{0, 2\}$-matching of the graph $(V, S)\times [V - \{u, v\}]\}$

$\cup\{\alpha - ax$: x is the incidence vector of a hamilton path in (V, S) from u to $v\}$.

[1b] Let $S \equiv S \cup \{j\}$.

End

Notice that sequential lifting leaves all old coefficients and the right-hand side of the inequality unchanged. The idea is to (sequentially) define each a_j as large as possible such that the inequality will remain valid, considering edges in $S \cup \{j\}$. Further, suppose we have a set X' of $|E'|$ affinely independent members x of $P(G')$ satisfying $ax = \alpha$. We can enlarge it to such a set X for $ax \leqslant \alpha$ and $P(G)$ by adding the following step:

[1c] For each $x \in X'$, add a new component $x_j \equiv 0$. If the minimum in [1a] was achieved by a $\{0, 2\}$-matching, let $\hat{x}^j$ be this $\{0, 2\}$-matching extended by defining $\hat{x}^j_j \equiv 2$. If the minimum was achieved by a hamilton path, let $\hat{x}^j$ be the tour obtained by defining $\hat{x}^j_j \equiv 1$. Let $X' \equiv X' \cup \{\hat{x}^j\}$.

Then, if we let X be the final X', the 'triangular structure' of the $\hat{x}^j$ will assure that X is affinely independent. This verifies that $ax \leqslant \alpha$ is indeed facet inducing for $P(G)$. This means, of course, that we will always finish with $a \geqslant 0$ (see Theorem 3.1).

Sequential lifting can be applied in many different orders to the edges, generally resulting in different lifted inequalities. Moreover, there can be facet inducing inequalities of $P(G)$ obtained by lifting from $a'x \leqslant \alpha'$, but not obtainable by sequential lifting.

Our main interest in this procedure is that it shows that 'unpleasant facets are, in effect, retained when edges are added to the graph. In particular, if we were to restrict our attention to complete graphs, Theorem 3.2 and Procedure 3.3 show that for n sufficiently large, there is a facet inducing inequality containing any desired consecutive pair of positive integers among the coefficients.

Finally, we can see, using Theorem 3.1, that the new coefficients defined by sequential lifting will never be larger than the largest previously existing coefficient, and generally, these new coefficients tend to decrease to zero as more edges are added.

We have the following conjecture.

Conjecture 3.4. *Let K_n be the complete graph on n nodes. For any positive integer s, there exists an integer $N(s)$ such that for $n \geqslant N(s)$, there is a facet inducing inequality of $P(K_n)$ whose coefficients include all integers from 0 to s.*

4. Facets of the travelling salesman polytope

In this section we discuss the relationship between facets of TSP(G) and facets of $P(G)$. Most studies of TSP(G) are restricted to the case of G being a complete graph, because solving travelling salesman problems on complete graphs is polynomially equivalent to the more general problem. An interesting feature of the results of the previous sections is that they do apply to general graphs. However, in this section we too will restrict ourselves to complete graphs to facilitate comparison with previously known results. We adopt the notation of Grötschel and Padberg [11] and let Q_T^n denote TSP(K_n). We let E_n and V_n respectively denote the edge set and node set of K_n. First we mention two preliminary results.

Proposition 4.1 ([11, Theorem 2.2]). *The dimension of Q_T^n is $n(n-3)/2 = |E_n| - |V_n|$ for $n \geq 3$.*

Corollary 4.2. *The minimal affine space containing Q_T^n is equal to*

$$\{x \in \mathbb{R}^{E_n}: x(\delta(i)) = 2 \text{ for all } i \in V_n\}.$$

The importance of this corollary is that it completely characterizes which inequalities induce the same facet as some prescribed facet inducing inequality for Q_T^n. We summarize this as follows.

Corollary 4.3. *Let $ax \leq \alpha$ be a valid inequality for Q_T^n. Then for any $\lambda = (\lambda_i \in \mathbb{R} : i \in V_n)$ and any $\mu > 0$, the inequality*

$$(\mu a)x + \sum_{i \in V_n} \lambda_i x(\delta(i)) \leq \mu\alpha + 2\lambda(V_n) \tag{4.1}$$

is a valid inequality for Q_T^n. Moreover, if we let $F \equiv \{x \in Q_T^n: ax = \alpha\}$, then the set of members of Q_T^n satisfying (4.1) *with equality is F. F is a facet if and only if every inequality whose corresponding hyperplane intersects Q_T^n in exactly F is of the form* (4.1) *for appropriate μ and λ.*

Of course, this corollary is a specialization of a fundamental polyhedral result: The inequalities in a linear system that defines a polyhedron are only unique up to positive multiples and the addition of equations satisfied by all members of the polyhedron.

For the remainder of this section, we let $P_n \equiv P(K_n)$, for $n \geq 3$. We now prove a basic result relating the facets of Q_T^n and P_n for $n \geq 3$, odd. It states that for each facet F of Q_T^n, there exists a *unique* facet F' of P_n such that $F = F' \cap Q_T^n$. This uniqueness is not true for general polyhedra, as illustrated in Fig. 4.1. In each case the 'ridge pole' is a face of the 'tent' and the marked end of the ridge pole is a facet of the face. Fig. 4.1(a) has the uniqueness property, but Fig. 4.1(b) does not.

Finally, we remark that the proof of the following theorem will consist of an

(a) (b)

Fig. 4.1.

algorithm which starts with a facet inducing inequality $ax \geqslant \alpha$ of Q_T^n and transforms it into a facet inducing inequality of P_n, which induces the same facet of Q_T^n.

Theorem 4.4. *For any facet F of Q_T^n, there exists a unique facet F' of P_n such that $F = F' \cap Q_T^n$.*

Proof. Let F be a facet of Q_T^n and suppose that $F = \{x \in Q_T^n\colon ax = \alpha\}$ where $ax \leqslant \alpha$ is a valid inequality for Q_T^n. We assume that $a \geqslant 0$. If not we add sufficiently high multiples of degree constraints so as to have this property. If we consider the inequality

$$ax + \sum_{i \in V} \lambda_i x(\delta(i)) \leqslant \alpha + 2 \sum_{i \in V} \lambda_i \tag{4.2}$$

we can see that varying λ_i for a node i has no effect on the feasibility of a $\{0, 2\}$-matching $\bar{x}$ satisfying $\bar{x}(\delta(i)) = 2$ or on a tour $\bar{x}$ which also must satisfy $\bar{x}(\delta(i)) = 2$. However, if $\bar{x}$ is a $\{0, 2\}$-matching deficient at i, then by choosing an appropriate value for λ_i, we can ensure

every $\{0, 2\}$-matching deficient at i satisfies (4.2), (4.3)

there exists an np-$\{0, 2\}$-matching deficient at i which satisfies (4.2) with equality. (4.4)

In fact it can be verified that this is given by

$$\lambda_i \equiv \tfrac{1}{2} \max\{ax - \alpha\colon x \text{ is a } \{0, 2\}\text{-matching of } G \text{ deficient at } i\}. \tag{4.5}$$

Since $a \geqslant 0$, this maximum will always be attained for a near perfect matching. Since the choice of deficient node provides a partition of the near perfect matchings, the λ_i are determined independently and uniquely. Thus, using Corollary 4.3, there is a unique (up to a positive multiple) inequality that is valid for Q_T^n and P_n, induces F and satisfies (4.3) and (4.4), two necessary conditions for it to be facet inducing for P_n. This is (4.2) with λ_i defined as in (4.5). Let $F' \equiv \{x \in P_n : x$ satisfies (4.2) with equality, for λ_i as in (4.5)$\}$. We show that F' is a facet of P_n.

Since F is a facet of Q_T^n, there exists a set T of $|E_n| - |V_n|$ affinely independent tours satisfying (4.2) with equality. For each node i, our choice of λ_i ensures that there exists a np-{0, 2}-matching x^i deficient at i satisfying (4.2) with equality. Let $M \equiv \{x^i : i \in V_n\}$. Then, for each $i \in V_n$, x^i is the only member of $T \cup M$ which does not satisfy $x(\delta(i)) = 2$. Thus it is affinely independent from $T \cup M - \{x^i\}$. Therefore $T \cup M$ is affinely independent of cardinality $|E_n|$ so F' is a facet of P_n and $F' \cap P_n = F$. □

Perhaps surprisingly, there are presently only three classes of facets for Q_T^n appearing in the literature. The first such class, the so called 'trivial' facets, are those induced by nonnegativity constraints $x_j \geq 0$ for all $j \in E_n$ [11, Theorem 3.2]. These obviously correspond to the inequalities (2.9) for P_n.

The second class of facets are those induced by the subtour elimination constraints $x(\gamma(S)) \leq |S| - 1$ for $S \subseteq V_n$, $2 \leq |S| \leq n - 1$ [11, Theorem 6.1]. For any such S, the subtour elimination constraints corresponding to S and $V_n - S$ induce the same facet of Q_T^n. (Simply sum one half the degree constraints for all nodes in $V_n - S$, subtract one half the sum of degree constraints for nodes in S and add this to the constraint $x(\gamma(S)) \leq |S| - 1$.) In particular, the edge 'capacity' constraints $x_j \leq 1$ for $j \in E$ induce the same facets as the subtour elimination constraints for the cardinality $n - 2$ subsets of V_n. By Theorem 4.4 there exists a unique facet of P_n which determines this 'doubly defined' facet of Q_T^n. Of course, this is the inequality (2.11) for the odd cardinality one of $|S|$, $|V_n - S|$.

The third class of facets, induced by generalized comb constraints is more complex. Let $W_i \subseteq V_n$ for $i = 0, 1, \ldots, k$ satisfy

$$|W_0 \cap W_i| \geq 1 \quad \text{for } i = 1, 2, \ldots, k, \tag{4.6}$$

$$|W_i - W_0| \geq 1 \quad \text{for } i = 1, 2, \ldots, k, \tag{4.7}$$

$$|W_i \cap W_j| = 0 \quad \text{for } 1 \leq i < j \leq k. \tag{4.8}$$

Then we call the graph C with nodeset $\bigcup_{i=0}^k W_i$ and edge set $\bigcup_{i=0}^k \gamma(W_i)$ a *comb* in K_n; W_0 is the *handle* and W_i are called the *teeth* for $i = 1, \ldots, k$. The comb inequality corresponding to C is given by

$$\sum_{i=0}^{k} x(\gamma(W_i)) \leq |W_0| + \sum_{i=1}^{k} (|W_i| - 1) - \lceil k/2 \rceil$$

where for $r \in \mathbb{R}$, $\lceil r \rceil$ denotes the smallest integer no less than r.

Note that the coefficients of a comb inequality will be 0, 1 or 2. Such inequalities were introduced by Chvátal [1] who required equality in (4.6), resulting in a (0–1)-inequality. We call such a comb *simple*. In a simple comb, each tooth has exactly one node in the handle. In a general comb, a tooth may have several nodes in the handle, and all edges joining these nodes have coefficient 2 in the inequality. A major result of Grötschel and Padberg [11, Theorem 6.2] is that for $k \geqslant 3$, odd, every comb inequality induces a facet of Q^n_T.

It is a routine matter to apply the procedure of the proof of Theorem 4.4 in order to find the corresponding facet inducing inequality for P_n. We illustrate this with the following theorem.

Theorem 4.5. *Let C be a simple comb having an odd number k of teeth such that $|V(C)|$ is even. If $|V(C)| \leqslant n - \lceil k/2 \rceil$, then the facet of P_n corresponding to the facet of Q^n_T, induced by the comb inequality for C is obtained by sequentially lifting a facet of the form of Theorem 3.2 for a subgraph G of K_n.*

Proof. Let $ax \leqslant \alpha$ be the comb inequality corresponding to C. Let G be the spanning subgraph of K_n whose edge set consists of $E(C)$ together with those edges having at least one end not in the comb. That is, we exclude those edges both of whose ends are in $V(C)$, but which are not in $E(C)$. We will show first that an application of the procedure of the proof of Theorem 4.4 obtains a facet inducing inequality for $P(G)$ of the form of Theorem 3.2.

We first compute the value λ_i for each $i \in V_n$, as given by (4.5). For $i \in V(C)$, the maximum value of ax for a $\{0, 2\}$-matching x deficient at i is $|V(C)| - 2$. For $i \in V_n - V(C)$, this maximum is $|V(C)|$. Therefore

$$\lambda_i = \begin{cases} 1/2\lceil k/2 \rceil - 1 & \text{for } i \in V(C), \\ 1/2\lceil k/2 \rceil & \text{for } i \in V_n - V(C). \end{cases}$$

When we use these values of λ_i in (4.2), we obtain the following new coefficients a'_j for each edge j.

$$a'_j = \begin{cases} \lceil k/2 \rceil - 1 & \text{for } j \in E(C), \\ \lceil k/2 \rceil - 1 & \text{for } j \in \delta(V(C)), \\ \lceil k/2 \rceil & \text{for } j \in \gamma(V_n - V(C)). \end{cases}$$

Now let $S \equiv V_n - V(C)$. Then $G[S]$ is a complete graph on an odd number of nodes, and so is hypomatchable and nonseparable. The graph $G \times S$ can also be easily checked to be hypomatchable and nonseparable. Because $|S| = n - |V(C)| \geqslant \lceil k/2 \rceil$, G is hamiltonian. Moreover, the segment number

$\tau(G[S]) = \tau(C) = \lceil k/2 \rceil$ so the coefficients a'_j are in fact given by

$$a'_j = \begin{cases} \tau(G[S]) & \text{for } j \in \gamma(S), \\ \tau(G[S]) - 1 & \text{for } j \in E(G) - \gamma(S). \end{cases}$$

These then are the coefficients of a facet of the form of Theorem 3.2 for G, so the resulting right-hand side, α', must equal

$$(\tau(G[S]) - 1)(n-1) + |S| - 1$$

as prescribed by the theorem.

Now let $a''x \leq \alpha''$ induce the facet of P_n corresponding to the comb inequality for C. By again using the procedure of the proof of Theorem 4.4, we see that

$$\alpha''_j = \begin{cases} a'_j & \text{for } j \in E(G), \\ \tau(G[S]) - 2 & \text{for } j \in E_n - E(G), \end{cases}$$

$$\alpha'' = \alpha'.$$

We will complete the proof by proving that the values $a''_j = \tau(G[S]) - 2$ for $j \in E_n - E(G)$ are those given by sequential lifting.

Suppose that we have sequentially lifted the coefficients for the edges of some (possibly empty) subset J of $E_n - E(G)$ and obtained the desired value. Let $j \in E_n - (J \cup E(G))$. Let u, v be the ends of j, let G^j denote the graph $(V_n, J \cup E(G))$, and let $\bar{G}^j$ denote $G^j[V_n - \{u, v\}]$. Then the maximum value of a $\{0, 2\}$-matching of $\bar{G}^j$ is $\alpha' - 2(\tau(G[S]) - 1)$. The maximum sum of the edge costs of a hamilton path in G^j from u to v is $\alpha' - (\tau(G[S]) - 2)$. Thus sequential lifting will define $a''_j \equiv (\tau(G[S]) - 2)$ and the proof now follows by induction. □

Fig. 4.2 illustrates a small example of this process. Let C be the ten node-five tooth comb of Fig. 4.2(a), in K_{13}. The comb inequality gives each edge the coefficient 1 and has $\alpha = 7$. The graph G of Fig. 4.2(b) consists of C, the subgraph induced by the set S of non-comb nodes and all edges joining these two parts. The procedure of Theorem 4.4 calculates $\lambda_i = \frac{1}{2}$ for the nodes i of C and $\lambda_i = \frac{3}{2}$ for the nodes i of S. Thus the coefficients a'_j are as indicated, 3 for edges of $\gamma(S)$ and 2 for all other edges and $\alpha' = 26$. Sequential lifting will then cause all edges of $E_{13} - E(G)$ to have the coefficient 1, which induces the facet of P_{13} corresponding to the facet of Q_T^{13} induced by C.

It is certainly possible to study the results of applying the procedure of Theorem 4.4 to combs having an odd number of nodes. In fact, this can be used

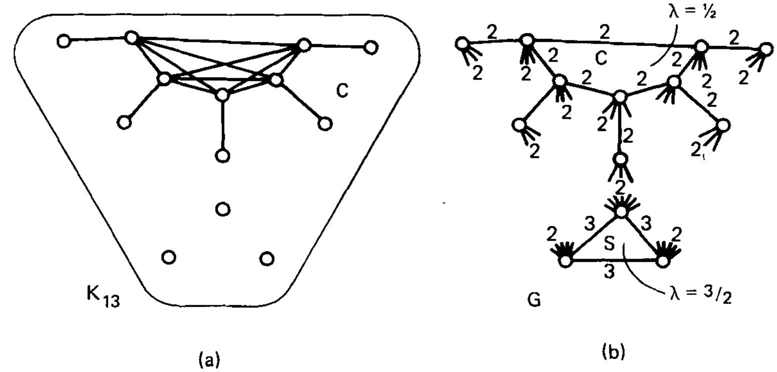


Fig. 4.2.

to provide other classes of non-(0–1) facets of P_n. However, for the remainder of this section we wish to briefly discuss cases when facet inducing inequalities of P_n induce facets of Q_T^n. In particular, for inequalities of the form (2.12), those induced by hypomatchable nonhamiltonian nonseparable edge maximal subgraphs. A graph G is said to be *hypohamiltonian* if G is nonhamiltonian, but $G-\{v\}$ is hamiltonian for all $v \in V$. It is an easy exercise to verify that if n is odd, then any edge maximal spanning hypohamiltonian subgraph of K_n satisfies our conditions of (2.12). Grötschel [9] showed that those spanning edge-maximal hypohamiltonian subgraphs of K_n which satisfy a certain technical property, do induce a facet of the monotone travelling salesman polytope. (He did not settle whether or not this technical property was indeed necessary.) Thus there is an obvious connection between our inequalities (2.12) and the monotone polytope. For the travelling salesman polytope itself, some inequalities (2.12) are facet inducing and some are not. For example the inequality (2.12) for the graph G_1 of Fig. 4.3(a) is facet inducing for Q_T^7, but that

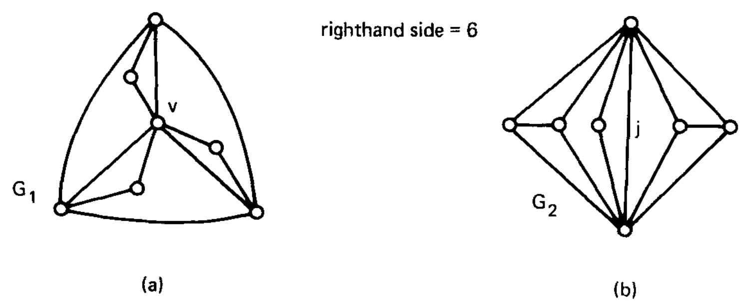


Fig. 4.3. (a) Facet inducing for Q_T^7; (b) Not facet inducing for Q_T^7.

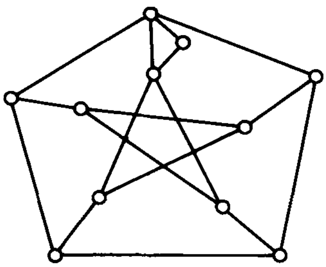


Fig. 4.4. Modified Petersen graph.

of the graph G_2 of Fig. 4.3(b) is not. (Note that both graphs satisfy the necessary conditions for (2.12) to apply, but neither is hypohamiltonian.)

The inequality (2.12) for G_1 is equivalent to the facet inducing inequality of the comb obtained by deleting node v. On the other hand, edge j of G_2 belongs to no hamilton cycle of K_7 which contains six edges of G_2. Therefore the inequality (2.12) for Q_T^7 is implied by the nonnegativity constraint for edge j.

However, there are examples of inequalities of the form (2.12) which are facet inducing for Q_T^n and which do not seem to arise from any known class of facet inducing inequalities. For example, the inequality (2.12) for the modified Petersen graph of Fig. 4.4 is facet inducing for K_{11}. This can be shown using a slight modification of the proof of Maurras [14] that the inequality $x(E) \leqslant 9$ is facet inducing for Q_T^{10}, where E is the edge set of any subgraph which is a Petersen graph.

5. Concluding remarks

When we wish to study a polytope such as the travelling salesman polytope, which is not of full dimension, we generally have considerable choice as to which full dimension polyhedron (if any) we will embed it in. We have studied, here, a particular polyhedron, P_n, which has several interesting properties. First (Theorem 4.4), for any facet of Q_T^n, there is a unique facet of P_n which intersects Q_T^n in exactly this facet. Thus we can partition the facets of P_n into three classes: those that contain all of Q_T^n, those that contain no facet of Q_T^n, and those that intersect Q_T^n in a facet. Theorem 4.4 shows that there is a bijection between the facets in the third class and the facets of Q_T^n.

In Theorem 2.12 we completely characterized those facets of $P(G)$, for general G, for which the inducing inequality can be scaled so as to have 0–1 valued coefficients. The most interesting set of facet inducing (0–1)-inequalities were those of (2.12). At the end of the previous section we saw that some of these do indeed induce facets of TSP(G) itself.

In Theorem 3.1, we determined several properties possessed by non-(0–1)-inequalities which induce facets of $P(G)$. One of these properties is that the subgraph of G induced by the edges having positive coefficients in such an inequality must be spanning, and indeed, must be hamiltonian. This has one rather negative consequence: Such inequalities will probably be harder to use in a cutting plane approach than, for example, the comb inequalities which have been used so successfully by Grötschel [10] to solve a 'real world' travelling salesman problem. However, a possible area for future research would be to see if 'simpler' equivalent inequalities (for TSP(G)) can be found for classes of such inequalities.

Acknowledgement

Much of this work was done when the authors were at I.M.A.G., Université Scientifique et Médicale de Grenoble, France, and the authors wish to acknowledge the support provided as well as the very stimulating environment. Moreover, the final version of this paper benefited substantially from the most helpful comments of Gilles Gastou and David Hartvigsen on an earlier version.

References

[1] V. Chvátal, 'Edmonds' polytopes and weakly Hamiltonian graphs, Math. Programming 5 (1973) 29–40.
[2] H.P. Crowder and M.W. Padberg, Solving large scale symmetric travelling salesman problems to optimality, Mgmt. Sci. 26 (1980) 495–509.
[3] G.B. Dantzig, D.R. Fulkerson and S.M. Johnson, Solution of a large-scale travelling salesman problem, Oper. Res. (1954) 393–410.
[4] J. Edmonds, Maximum matching and a polyhedron with 0, 1 vertices, J. Res. Nat. Bur. Standards Sect. B 69 (1965) 125–130.
[5] J. Edmonds, Matroids and the greedy algorithm, Math. Programming 1 (1971) 127–136.
[6] J. Edmonds, Matroid Intersection, Annals of Discrete Mathematics 4 (1979) 39–49.
[7] J. Edmonds, Optimum branchings, in: G.B. Dantzig and A.F. Veinott, eds., Mathematics of the Decision Sciences, Part 1 (American Mathematical Society, Providence, 1968) pp. 346–361.
[8] M. Grötschel, Polyedrische Characterisierungen kombinatorischer Optimierungsprobleme, Dissertation, Universität Bonn, 1977.
[9] M. Grötschel, On the monotone symmetric travelling salesman problem: hypohamiltonian/hypotraceable graphs and facets, Math. Oper. Res. 5 (1980) 285–292.
[10] M. Grötschel, On the symmetric travelling salesman problem: solution of a 120-city problem, Math. Programming Stud. 12 (1980) 61–77.
[11] M. Grötschel and M.W. Padberg, On the symmetric travelling salesman problem I, II, Math. Programming 16 (1979) 265–302.
[12] L. Lovász, A note on factor-critical graphs, Studia Sci. Math. Hungar. 7 (1972) 297–280.
[13] L. Lovász, Graph theory and integer programming, Ann. Discrete Math. 4 (1979) 141–158.
[14] J.F. Maurras, Polytopes à sommets dans $\{0, 1\}^n$, Dissertation, Univ. Paris VII, 1976.

[15] M.W. Padberg and S. Hong, On the symmetric travelling salesman problem: A computational study, T.J. Watson Res. Rept., IBM Research, Yorktown Heights, 1977.
[16] W.R. Pulleyblank and J. Edmonds, Facets of 1-matching polyhedra, in: C. Berge and D.K. Ray-Chaudhuri, eds., Hypergraph Seminar (Springer, Berlin, 1974) pp. 214–242.
[17] W.T. Tutte, The factorization of linear graphs, J. London Math. Soc. 22 (1947) 107–111.

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 57–67

POLYHEDRA FOR COMPOSED INDEPENDENCE SYSTEMS

William H. CUNNINGHAM*

Institute for Operations Research, University of Bonn, 5300 Bonn 1, West Germany

We consider a composition for independence systems, and show how knowledge of polyhedral descriptions of the small systems provides such a description for the composed system. Applications to other compositions, including sums and substitutions, are treated. Theorems of Chvátal on graph substitution and Bixby on the composition of perfect graphs are proved.

1. Introduction

An *independence system* H is a pair $(E, \mathcal{I})$ where E is a finite set and $\mathcal{I}$ is a non-empty family of subsets of E, called *independent* sets, satisfying

$$\text{If } A \subseteq B \in \mathcal{I}, \quad \text{then } A \in \mathcal{I}.$$

The *optimum independent set* problem for H is: Given a real weight c_j for each $j \in E$, find $I \in \mathcal{I}$ so that $\Sigma\,(c_j : j \in I)$ is maximized.

We remind the reader of two prototype examples of independence systems. Let G be a simple graph, let $E = V(G)$, and let $I \in \mathcal{I}$ if and only if I is a *stable* set of G, that is, no two elements of I are adjacent in G. We call this first example a graphic independence system. The optimum independent set problem for graphic systems is well known to be NP-hard in general. It can be solved in polynomial time for a number of special classes of graphs, for example, bipartite graphs, edge graphs, perfect graphs, and claw-free graphs.

The second prototype example occurs when E is the set of elements and $\mathcal{I}$ is the family of independent sets of a *matroid*. These are defined axiomatically as independence systems which also satisfy:

For every $A \subseteq E$, all maximal independent subsets of A have the same cardinality.

* Supported by Sonderforschungsbereich 21 (DFG), Institut für Ökonometrie und Operations Research, Universität Bonn, W. Germany. On leave from Department of Mathematics and Statistics, Carleton University. Research partially supported by a grant from N.S.E.R.C. of Canada.

An important special class of matroids consists of those for which E is the edge-set of a graph G and $\mathscr{I}$ is the set of edge-sets of forests of G. Another occurs when E indexes the columns of a matrix and 'independent' means 'linearly independent'. For matroids the optimum independent set problem can be solved by a simple 'greedy' algorithm, and this algorithm runs in polynomial time provided that there exists a similarly efficient algorithm for determining, for a given $A \subseteq E$, whether $A \in \mathscr{I}$.

For most of the classes of independence systems H for which the optimum independent set problem is well solved, there is also a polyhedral description of the independent sets. That is, there is an explicit (though possibly large) collection of inequalities whose solution set forms the convex hull P of (incidence vectors of) independent sets of H. (For example this is the case when H is a matroid, or when H is a graphic independence system arising from an edge graph or a perfect graph.) In this paper we investigate a certain composition for independence systems, introduced in [4], from a polyhedral viewpoint. We show how knowledge of explicit polyhedral descriptions for two independence systems leads to such a description for their composition. This result is applied to provide similar results for some other compositions; for example, we obtain a proof of the theorem of Chvátal [2] on 'graph substitution' which motivated this research.

2. Preliminaries

We will deal with polyhedral representations in the form

$$\sum (a_{ij}x_j : j \in E) \leq b_i, \quad i \in L, \tag{1}$$

$$x_j \geq 0, \quad j \in E. \tag{2}$$

The inequalities (2) are called *trivial*; those of (1) are *non-trivial*. If the solution set of (1), (2) is to be the convex hull P of independent sets of H, we can assume that $a_{ij} \geq 0$ for $i \in L$, $j \in E$; it will often be convenient also to assume that each $b_i = 1$. The latter assumption is valid provided that H has no 'loops', that is, no elements j with $\{j\} \notin \mathscr{I}$. (Of course, in treating the optimum independent set problem, loops would simply be ignored.) We will make this assumption. It is equivalent to the assumption that P has affine rank $|E|+1$ ('full dimension'). Also, in this situation each of the trivial inequalities defines a facet of P.

Two simple unary operations on independence systems are deletion and contraction. Given $H = (E, \mathscr{I})$ and an element $e \in E$, $H \backslash e$, the independence system obtained by *deleting* e, is defined to be $(E\backslash\{e\}, \mathscr{I}')$, where $\mathscr{I}' = \{I: E\backslash\{e\} \supseteq I \in \mathscr{I}\}$; H/e, the independence system obtained from H by *contracting* e is defined to be $(E\backslash\{e\}, \mathscr{I}'')$, where $\mathscr{I}'' = \{I: I \subseteq E\backslash\{e\}, I \cup \{e\} \in \mathscr{I}\}$. (Notice that contracting an element may introduce loops; these may then be deleted.)

Proposition 1. *Let $H = (E, \mathscr{I})$ be an independence system and let $e \in E$. If non-trivial inequalities for H are*

$$\sum (a_{ij}x_j: j \in E) \leq 1, \quad i \in L,$$

then non-trivial inequalities for $H\backslash e$ are

$$\sum (a_{ij}x_j: j \in E\backslash\{e\}) \leq 1, \quad i \in L,$$

and non-trivial inequalities for H/e are

$$\sum (a_{ij}x_j: j \in E\backslash\{e\}) \leq 1 - a_{ie}, \quad i \in L.$$

Proof. Let P' be the polyhedron for $H\backslash e$. Then if $x = (x_j: j \in E\backslash\{e\})$ is in P', clearly $(0, x) \in P$. But the converse is also true, because if $(0, x) \in P$, then $(0, x)$ is a convex combination of independent sets of H, none of which contains e. So $x \in P'$ if and only if $(0, x) \in P$, which is true if and only if ($x \geq 0$, and) $\Sigma\,(a_{ij}x_j: j \in E\backslash\{e\}) \leq 1$, $i \in L$, as required.

Now let P'' be the polyhedron for H/e. If $x = (x_j: j \in E\backslash\{e\})$ is in P'', then $(1, x) \in P$. Again the converse is true, because if $(1, x) \in P$, then $(1, x)$ is a convex combination of independent sets each of which contains e, so x is a convex combination of independent sets of H/e. Thus $x \in P''$ if and only if $(1, x) \in P$, and so $x \in P''$ if and only if ($x \geq 0$, and) $a_{ie} + \Sigma\,(a_{ij}x_j: j \in E\backslash\{e\}) \leq 1$, as required. □

It is worth remarking that, in addition to being able to get very easily polyhedral descriptions for $H\backslash e$ and H/e from one for H, we also can solve the optimum independent set problem for $H\backslash e$ and H/e, whenever we can solve it for H. In the first case, we assign e a sufficiently large negative weight, and in the second case a large positive weight, and then solve the problem on H.

We now introduce an elementary composition for independence systems. Let $H_1 = (E, \mathscr{I}_1)$ and $H_2 = (E_2, \mathscr{I}_2)$ be independence systems for which $E_1 \cap E_2 = \emptyset$.

The *direct sum* of H_1 and H_2 is $H=(E,\mathscr{I})$ where $E=E_1\cup E_2$ and $\mathscr{I}=\{I_1\cup I_2\colon I_1\in\mathscr{I}_1, I_2\in\mathscr{I}_2\}$. The following result is easy to prove.

Proposition 2. *For $k=1$ and 2, let non-trivial inequalities for H_k be*

$$\sum(a^k_{ij}x_j\colon j\in E_k)\leq 1,\quad i\in L_k.$$

If H is the direct sum of H_1 and H_2, then non-trivial inequalities for H are

$$\sum(a^1_{ij}x_j\colon j\in E_1)\leq 1,\quad i\in L_1,$$

$$\sum(a^2_{ij}x_j\colon j\in E_2)\leq 1,\quad i\in L_2.$$

An independence system is said to be *separable* if it has a non-trivial expression (one in which $E_1\neq\emptyset\neq E_2$) as a direct sum. Since these systems can be handled in a straightforward manner by restricting attention to their irreducible summands (this also applies to the optimum independent set problem), we henceforth assume that independence systems are non-separable. For technical reasons we also assume that each has at least two elements. These two assumptions already imply that there are no loops. We point out that the direct sum has a very natural interpretation as applied to graphic independence systems, or to matroids arising from graphs or matrices.

3. A composition theorem

The main results of this paper concern the composition which we define now. Let $H_1=(E_1\cup\{e\},\mathscr{I}_1)$ and $H_2=(E_2\cup\{e\},\mathscr{I}_2)$ be independence systems, where $E_1\cap E_2=\emptyset$ and $e\notin E_1\cup E_2$. The *composition* of H_1 and H_2 is the independence system $H_1*H_2=(E,\mathscr{I})$ where $E=E_1\cup E_2$ and $\mathscr{I}=\{I_1\cup I_2\colon I_1\in\mathscr{I}_1$ and $I_2\cup\{e\}\in\mathscr{I}_2$, or $I_1\cup\{e\}\in\mathscr{I}_1$ and $I_2\in\mathscr{I}_2\}$. This composition was introduced in [4], and was treated mainly from the viewpoint of 'circuits'-minimal non-independent sets. It is not difficult to show that, if H_1 and H_2 are non-separable and have at least two elements, then H has these properties too. In the case in which H_1 and H_2 are graphic, arising from connected graphs G_1, G_2 respectively, H_1*H_2 is also graphic, and arises from the graph G constructed as follows. Delete the vertex e from G_1 and G_2 and join every neighbour of e in G_1 to every neighbour of e in G_2. Similarly, if H_1 and H_2 are matroids, then so is H_1*H_2 (see [4]). Moreover, if H_1 and H_2 are forest matroids of graphs G_1, G_2, then H_1*H_2 is the forest matroid of a graph G arising from G_1, G_2 as follows: Identify the edge e of G_1 with the edge e of G_2, and then delete e.

Before describing the construction of a polyhedron for $H_1 * H_2$ from those for H_1 and H_2, we make an observation relevant to the optimum independent set problem for $H_1 * H_2$. Every independent set of $H_1 * H_2$ is of the form $I_1 \cup I_2$ where either I_1 is independent in $H_1\backslash e$ and I_2 is independent in H_2/e or I_1 is independent in H_1/e and I_2 is independent in $H_2\backslash e$. Therefore one way to solve the independent set problem over $H_1 * H_2$ is: Solve the problem over $H_2\backslash e$ (with the given weights c_j) to obtain I_2 and over H_2/e to obtain I_2'. Then assign e weight $\Sigma\,(c_j\colon j \in I_2) - \Sigma\,(c_j\colon j \in I_2')$ and solve the problem over H_1 to obtain I_1. Then an optimal independent set for $H_1 * H_2$ is $(I_1 \cup I_2)\backslash\{e\}$ if $e \in I_1$ and $I_1 \cup I_2'$ otherwise. The above construction was described for the case of graphic independence systems in [3].

The main polyhedral result of the paper is the following.

Theorem 1. *For $k = 1$ and 2 let non-trivial inequalities for $H_k = (E_k \cup \{e\}, \mathcal{I}_k)$ be*

$$\sum (a_{ij}^k x_j\colon j \in E_k \cup \{e\}) \leq 1, \quad i \in L_k. \tag{3}$$

Then non-trivial inequalities for $H_1 * H_2$ are

$$a_{me}^2 \sum (a_{ij}^1 x_j\colon j \in E_1) + a_{ie}^1 \sum (a_{mj}^2 x_j\colon j \in E_2) \leq a_{ie}^1 + a_{me}^2 - a_{ie}^1 a_{me}^2, \quad i \in L_1,\ m \in L_2. \tag{4}$$

Proof. Let P, P_1, P_2 be the polyhedra for $H_1 * H_2$, H_1, H_2 respectively. We claim that, where x^k denotes $(x_j\colon j \in E_k)$ for $k = 1$ and 2, then $(x^1, x^2) \in P$ if and only if there exists a number α, $0 \leq \alpha \leq 1$, such that $(\alpha, x^1) \in P_1$ and $(1 - \alpha, x^2) \in P_2$. Suppose that this claim is true. Then, to find inequalities for P, we can use Fourier elimination to eliminate α from the system:

$$a_{ie}\alpha + \sum (a_{ij}^1 x_j\colon j \in E_1) \leq 1, \quad i \in L_1, \tag{5}$$

$$\begin{gathered} a_{me}^2(1 - \alpha) + \sum (a_{mj}^2 x_j\colon j \in E_2) \leq 1, \quad m \in L_2, \\ x_j \geq 0, \quad j \in E_1 \cup E_2, \qquad 0 \leq \alpha \leq 1. \end{gathered} \tag{6}$$

The inequalities obtained from adding a_{me}^2 times an inequality from (5) to a_{ie}^1 times an inequality from (6) are precisely the inequalities of (4). The other inequalities arising from this process are ($x_j \geq 0$, $j \in E_1 \cup E_2$, and) inequalities of the form

$$\sum (a_{ij}^1 x_j\colon j \in E_1) \leq 1 - a_{ie}^1, \quad i \in L_1 \text{ (and similarly for } L_2\text{)}.$$

But it is easy to see that there must be $m \in L_2$ for which $a_{me}^2 = 1$; choosing i

and this m in (4) leads to an inequality at least as strong as

$$\sum (a^1_{ij}x_j : j \in E_1) \leqslant 1 - a^1_{ie}\,.$$

Now we must prove the claim. Suppose first that $x = (x^1, x^2) \in P$. Then $x = \Sigma\, \lambda_i(I^2_i \cup J^1_i) + \Sigma\, \mu_j(I^1_j \cup J^2_j)$ where λ_i, $\mu_j \geqslant 0$, $\Sigma\, \lambda_i + \Sigma\, \mu_j = 1$, $I^1_i \in \mathscr{I}_1$, $J^1_i \cup \{e\} \in \mathscr{I}_1$, $I^2_j \in \mathscr{I}_2$, $J^2_j \cup \{e\} \in \mathscr{I}_2$. Take $\Sigma\, \lambda_i = \alpha$. Then $\Sigma\, \lambda_i(J^1_i \cup \{e\}) + \Sigma\, \mu_j I^1_j = (\alpha, x^1) \in P_1$. Similarly $(1-\alpha, x^2) \in P_2$.

Now suppose that $(\alpha, x^1) \in P_1$ and $(1-\alpha, x^2) \in P_2$ for some α, $0 \leqslant \alpha \leqslant 1$. Then $(\alpha, x^1) = \Sigma\, \lambda^1_i I^1_i + \Sigma\, \mu^1_k J^1_k$ and $(1-\alpha, x^2) = \Sigma\, \lambda^2_j I^2_j + \Sigma\, \mu^2_m J^2_m$, where λ^1_i, μ^1_k, λ^2_j, $\mu^2_m \geqslant 0$, $\Sigma\, \lambda^1_i + \Sigma\, \mu^1_k = \Sigma\, \lambda^2_j + \Sigma\, \mu^2_m = 1$, $e \notin I^1_i \in \mathscr{I}_1$, $e \in J^1_k \in \mathscr{I}_1$, $e \notin I^2_j \in \mathscr{I}_2$, $e \in J^2_m \in \mathscr{I}_2$. Then $(I^1_i \cup J^2_m) \backslash \{e\} \in \mathscr{I}$ for all i and m, and $(I^2_j \cup J^1_k) \backslash \{e\} \in \mathscr{I}$ for all j and k, and it is straightforward to get (x^1, x^2) as a convex combination of these. This completes the proof. □

When considering polyhedral descriptions for independence systems, it is natural to be interested in minimal descriptions, that is, descriptions for which every inequality is essential to the definition of the polyhedron. Equivalently, since our polyhedra have full dimension, we are interested in knowing which of the inequalities in a description of the system correspond to facets of the polyhedron. In the context of Theorem 1, if both of the inequalities from (3) used to create an inequality from (4) correspond to facets of their polyhedra, then the resulting inequality (4) usually yields a facet too. This can fail when one of the original inequalities is $x_e \leqslant 1$; it may be that this is essentially the only way it can fail, but I have been unable to prove this. There are special cases in which one can prove that all of the resulting inequalities must yield facets if the original ones did; an example is the substitution composition for graphs described in the next section.

4. Substitution composition

There are a number of applications of Theorem 1 to the proof of polyhedral results on other kinds of compositions. As an example, consider the following composition for graphs, called *substitution* by Chvátal [2]. Let G_1, G_2 be graphs having disjoint vertex sets and let e be a vertex of G_1. The graph G obtained by substituting G_2 for the vertex e of G_1 is obtained as follows. Delete e from G_1, and join every neighbour of e in G_1 to every vertex of G_2. We can generalize this graph composition to a composition of independence systems, in the following way. Let $H_1 = (E_1 \cup \{e\}, \mathscr{I}_1)$ and $H_2 = (E_2, \mathscr{I}_2)$ be independence systems with $E_2 \cap (E_1 \cup \{e\}) = \emptyset$. Let $H = (E, \mathscr{I})$ be defined by $E = E_1 \cup E_2$ and $\mathscr{I} = \{I: e \notin I \in \mathscr{I}_1\} \cup \{(I_1 \cup I_2)\backslash\{e\}: e \in I_1 \in \mathscr{I}_1, I_2 \in \mathscr{I}_2\}$; we say that H is obtained by *substituting* H_2 for the element e of H_1, and write $H = H_1[H_2; e]$. For the special case of graph substitution, the following is a theorem of Chvátal [2], which motivated the present work.

Theorem 2. *Let*

$$\sum (a^1_{ij}x_j: j \in E_1 \cup \{e\}) \leq 1, \quad i \in L_1, \tag{7}$$

$$\sum (a^2_{mj}x_j: j \in E_2) \leq 1, \quad m \in L_2, \tag{8}$$

be non-trivial inequalities for $H_1 = (E_1 \cup \{e\}, \mathscr{I}_1)$ *and* $H_2 = (E_2, \mathscr{I}_2)$ *respectively. Then non-trivial inequalities for* $H_1[H_2; e]$ *are*

$$\sum (a^1_{ij}x_j: j \in E_1) + a^1_{ie} \sum (a^2_{mj}x_j: j \in E_2) \leq 1, \quad i \in L_1,\ m \in L_2. \tag{9}$$

Proof. Define H_2' to be $(E_2 \cup \{e\}, \mathscr{I}_2 \cup \{e\})$. (In the graph case, this amounts to adding e as a vertex to G_2 and joining it to every other vertex.) Then it is easy to see that $H_1 * H_2' = H_1[H_2; e]$. Therefore, if we can determine non-trivial inequalities for H_2' from those for H_2, we can just apply Theorem 1. We claim that such inequalities are

$$x_e + \sum (a^2_{ij}x_j: j \in E_2) \leq 1, \quad i \in L_2. \tag{10}$$

The proof is easy: where $x^2 = (x_j: j \in E_2)$, (x_e, x^2) is in the polyhedron for H_2' if and only if ($0 \leq x_e \leq 1$, and) $(1 - x_e)x^2 \in P_2$, which is true if and only if (10) holds, as required. Now applying Theorem 1 to (7) and (10) gives (9), and the proof is complete. $\square$

(There is a second method of substituting an independence system $H_2 = (E_2, \mathscr{I}_2)$ for an element e of an independence system $H_1 = (E_1 \cup \{e\}, \mathscr{I}_1)$. The resulting independence system $H = (E, \mathscr{I})$ is defined as follows: $E = E_1 \cup E_2$ and $\mathscr{I} = \{I_1 \cup I_2: I_1 \cup \{e\} \in \mathscr{I}_1 \text{ or } I_2 \in \mathscr{I}_2\}$; it is treated in [4], for example. This composition can also be seen to generalize the substitution composition for graphs, when we change the definition of 'independent' to "does not contain the vertex-set of a clique". Another motivation for considering this composition is its relation to a natural method of composing Boolean functions. An independence system is associated with a monotone Boolean function f as follows: $f(A) = 0$ if and only if A is independent. The composition referred to arises from substituting the values of one Boolean function into a variable of another. This second substitution composition can also be derived from the $*$-composition. If we form $H_2'' = (E_2 \cup \{e\}, \mathcal{I}_2'')$ by defining $\mathcal{I}_2'' = \{I : e \not\in I\} \cup \{I \cup \{e\} : I \in \mathcal{I}_2\}$, then $H = H_1 * H_2''$. Therefore, a result like Theorem 2 for this composition can be derived from Theorem 1. We leave the details to the reader.)

5. Perfect and semi-perfect systems

An independence system $H = (E, \mathcal{I})$ is *perfect* if and only if there is a collection of non-trivial inequalities for H having $b_i = 1$ and $a_{ij} = 0$ or 1 for all i. It is well known (see [2], for example) that an independence system is perfect if and only if it arises from the stable sets of a perfect graph. It is easy to see from Theorem 2 that, if H_1, H_2 are perfect, then so is $H_1[H_2; e]$, so Chvátal's theorem generalizes the result that the substitution composition of perfect graphs is again perfect. Similarly, one can see from Theorem 1 that if H_1 and H_2 are perfect then so is $H_1 * H_2$. Therefore, Theorem 2 generalizes a theorem of Bixby [1], that the $*$-composition of perfect graphs is perfect.

Let us call an independence system $H = (E, \mathcal{I})$ *semi-perfect* if there is a collection of non-trivial inequalities for H such that $b_i = 1$ and $a_{ij} = 0$ or ε_i for all i. Clearly, perfect independence systems are semi-perfect; so are matroids and graphic systems arising from edge-graphs (matchings). It is easy to see from Theorem 1 that if H_1 and H_2 are semi-perfect, then so is $H_1 * H_2$. Thus, for example, if H_1 and H_2 are matroids, then $H_1 * H_2$ is semi-perfect; but this of course follows from the fact [4] that $H_1 * H_2$ is itself a matroid. Interestingly, the substitution composition does not preserve semi-perfection.

6. Sums

The direct sum composition is a special case of a general construction for independence systems. Given systems $H_1 = (E_1, \mathcal{I}_1)$ and $H_2 = (E_2, \mathcal{I}_2)$, their *sum* $H_1 + H_2 = (E, \mathcal{I})$ is defined by $E = E_1 \cup E_2$ and $\mathcal{I} = \{I_1 \cup I_2 : I_1 \in \mathcal{I}_1, I_2 \in \mathcal{I}_2\}$. The direct sum is the special case in which $E_1 \cap E_2 = \emptyset$. For the more general sum it would be useful to be able to obtain an explicit description of the polyhedron for $H_1 + H_2$ in terms of such descriptions for H_1, H_2. When H_1, H_2 are matroids, $H_1 + H_2$ is also a matroid, and there is a formula [6] for its rank function in terms of the rank functions of H_1, H_2. Therefore, in this special case such a description of the polyhedron for H_1+H_2 is available.

In general, however, we should not expect such a description for the polyhedron of the sum to be easy to discover, as the following example shows. Let G be a 3-regular graph having edge-set E, and let $H=(E,\mathscr{I})$, where $\mathscr{I}=\{I\subseteq E: I$ is a matching of $G\}$. Then $H'=H+H$ is an independence system $(E,\mathscr{I}')$, and there exist $I\in\mathscr{I}'$ with $|I|=|V|$ if and only if G is 3-edge colourable. On the one hand, there exists an explicit description for the polyhedron of H [5]; on the other hand, the problem of deciding whether such $\mathscr{I}$ exists is NP-hard [7].

However, we can provide a description of the sum in another special case, namely when H_1, H_2 have just one common element, as an application of Theorem 1. In this special case, the sum is sometimes called the 'series connection' of H_1 and H_2. The polyhedral description for this composition is given in the following result.

Theorem 3. *Let $H_k=(E_k\cup\{e\},\mathscr{I}_k)$ for $k=1$ and 2 be independence systems, where $E_1\cap E_2=\emptyset$, and suppose that a collection of non-trivial inequalities for H_k is*

$$\sum (a^k_{ij}x_j: j\in E_k\cup\{e\})\leqslant 1,\quad i\in L_k. \tag{11}$$

Then a collection of non-trivial inequalities for H_1+H_2 is

$$x_e\leqslant 1, \tag{12}$$

$$\sum (a^k_{ij}x_j: j\in E_k)\leqslant 1,\quad i\in L_k,\ k=1 \text{ and } 2, \tag{13}$$

$$a^1_{ie}a^2_{me}x_e+a^2_{me}\sum (a^1_{ij}x_j: j\in E_1)+a^1_{ie}\sum (a^2_{mj}x_j: j\in E_2)\leqslant a^1_{ie}+a^2_{me},$$
$$i\in L_1,\ m\in L_2. \tag{14}$$

Proof. The proof is based on the existence of an expression for H_1+H_2 in terms of $*$-compositions. (There is a sort of converse, that $H_1*H_2=(H_1+H_2)/e$, but we will not need this.) Let H_1' be obtained from H_1 by replacing e by f and H_2' be obtained from H_2 by replacing e by g, where $f, g\notin E_1\cup E_2\cup\{e\}$. Let $H'=(E',\mathscr{I}')$ be defined by $E'=\{e,f,g\}$ and $\mathscr{I}'=\{I\subseteq E': |I|<3\}$. Then it is straightforward to verify that $H_1+H_2=H_1'*H'*H_2'$.

Now, we apply Theorem 1 to $H_1' * H' * H_2'$. A collection of non-trivial inequalities for H' is clearly

$$x_e \leqslant 1\,, \qquad x_f \leqslant 1\,, \qquad x_g \leqslant 1\,, \qquad \tfrac{1}{2}x_e + \tfrac{1}{2}x_f + \tfrac{1}{2}x_g \leqslant 1\,.$$

Therefore, by Theorem 1, a list of non-trivial inequalities for $H_1' * H'$ is

$$x_e \leqslant 1\,, \qquad x_g \leqslant 1\,, \tag{15}$$

$$\sum (a_{ij}^1 x_j : j \in E_1) \leqslant 1, \quad i \in L_1\,, \tag{16}$$

$$\tfrac{1}{2}\sum (a_{ij}^1 x_j : j \in E_1) + a_{ie}^1(\tfrac{1}{2}x_e + \tfrac{1}{2}x_g) \leqslant \tfrac{1}{2} + a_{ie}^1 - \tfrac{1}{2}a_{ie}^1, \quad i \in L_1\,. \tag{17}$$

Simplifying, we get (15), (16) and

$$\frac{1}{1 + a_{ie}^1}\left[\sum (a_{ij}^1 x_j : j \in E_1) + a_{ie}^1 x_e + a_{ie}^1 x_g\right] \leqslant 1, \quad i \in L_1\,. \tag{18}$$

Applying Theorem 1 again to compose $H_1' * H'$ with H_2', we obtain

$$x_e \leqslant 1\,,$$

$$\sum (a_{ij}^1 x_j : j \in E_1) \leqslant 1, \quad i \in L_1\,,$$

$$\sum (a_{mj}^2 x_j : j \in E_2) \leqslant 1, \quad m \in L_2\,,$$

$$a_{me}^2 \frac{a_{ie}^1}{1 + a_{ie}^1} x_e + a_{me}^2 \sum \left(\frac{a_{ij}^1}{1 + a_{ie}^1} x_j : j \in E_1\right) + \frac{a_{ie}^1}{1 + a_{ie}^1} \sum (a_{mj}^2 x_j : j \in E_2) \leqslant$$
$$\leqslant a_{me}^2 + \frac{a_{ie}^1}{1 + a_{ie}^1} - \frac{a_{me}^2 a_{ie}^1}{1 + a_{ie}^1}, \quad i \in L_1, m \in L_2\,.$$

Multiplying the last inequalities by $1 + a_{ie}^1$ and simplifying, we have the desired collection of inequalities, and the proof is complete. □

References

[1] R.E. Bixby, private communication, 1975.
[2] V. Chvátal, On certain polytopes related to graphs, J. Combin. Theory Ser. B 18 (1975) 138–154.

[3] W.H. Cunningham, Decomposition of directed graphs, SIAM J. Algebraic and Discrete Methods, to appear.

[4] W.H. Cunningham and J. Edmonds, A combinatorial decomposition theory, Canad. J. Math. 32 (1980) 734–765.

[5] J. Edmonds, Maximum matching and a polyhedron with 0–1 vertices, J. Res. Nat. Bur. Stand. B 69 (1965) 125–130.

[6] J. Edmonds and D.R. Fulkerson, Transversals and matroid partition, J. Res. Nat. Bur. Stand. B 69 (1965) 147–153.

[7] I. Holyer, The NP-completeness of edge-colouring, SIAM J. Comp. 10 (1981) 718–720.

Annals of Discrete Mathematics 16 (1982) 69–82

AUGMENTING PATHS AND A CLASS OF INDEPENDENCE SYSTEMS

Reinhardt EULER

Mathematisches Institut der Universität zu Köln, Weyertal 86–90, D-5 Köln 41, W. Germany

We generalize independence systems $(E, \mathcal{I})$ arising from claw-free graphs, 2-matroid intersections and b-matchings by weakening matroid axioms. Motivated by the augmenting path theory for claw-free graphs (cf. [8, 10]) as well as 2-matroid intersections (cf. [6]) we introduce the concept of the Dependence Graph DPG(Z, X) where X is supposed to be an independent set and $Z \subseteq E\backslash X$. If $X, Y \in \mathcal{I}$ such that $|Y| = |X| + 1$, we show that a so-called W-path W always exists in DPG$(Y\backslash X, X)$. As $(X\backslash W) \cup (W\backslash X)$ does not need to be independent, implications for this exchange as well as for an augmentation from X to an independent set X' having one element more than X are discussed.

1. Introduction

Given a finite set E an independence system $(E, \mathcal{I})$ consists of a nonempty system $\mathcal{I}$ of subsets of E satisfying

$$Y \subseteq X \in \mathcal{I} \Rightarrow Y \in \mathcal{I}.$$

We call a set $X \subseteq E$ *independent* if $X \in \mathcal{I}$, *dependent* otherwise. A *base of* X, $X \subseteq E$, is a maximal independent subset of X, a *circuit* of $(E, \mathcal{I})$ is a minimal dependent subset of E (maximal and minimal with respect to set inclusion). An independent set of maximal cardinality will also be called a *maximum* independent set. Without loss of generality we can restrict ourselves to independence systems, which are *normal*, i.e., $\{e\} \in \mathcal{I}$ for all $e \in E$.

If $\mathcal{A} = (A_i, i \in I)$ is a family of subsets of E, then $T \subseteq E$ is called a *transversal* of $\mathcal{A}$, if there exists a bijection $\pi : T \rightarrow I$ such that

$$x \in A_{\pi(x)} \quad \text{for all } x \in T.$$

An extension of Hall's theorem for the existence of a transversal of $\mathcal{A}$ (cf. [9]) says that the family $\mathcal{A}$ can be partitioned into r subfamilies each of which possesses a transversal, if and only if

$$|K| \leqslant r \left| \bigcup_{k \in K} A_k \right| \quad \text{for all } K \subseteq I.$$

Let $(E, \mathcal{I})$ be an independence system and $X \in \mathcal{I}$, $Z \subseteq E$. Then for each

$e \in E \backslash X$ we denote by $\mathscr{C}(e, X)$ the system of all circuits contained in $X \cup \{e\}$ and by $\mathscr{C}(Z, X)$ the system $\bigcup_{e \in Z \backslash X} \mathscr{C}(e, X)$. The families considered here for $X, Y \in \mathscr{I}$ are of the type $(C \backslash Y, C \in \mathscr{C}(Y, X))$.

2. (2, 2)-Systems and examples

Definition 2.1. Let $(E, \mathscr{I})$ be an independence system and p, r be the minimal, positive integers such that

(i) for all $X \in \mathscr{I}$, $e \in E \backslash X$ the set $X \cup \{e\}$ contains at most p distinct circuits;

(ii) for all $X, Y \in \mathscr{I}$ the family $(C \backslash Y, C \in \mathscr{C}(Y, X))$ can be partitioned into r subfamilies each of which possesses a transversal.

Then $(E, \mathscr{I})$ is called a *(p, r)-system.*

The motivation for introducing the concept of (p, r)-systems is to describe two, in general not equivalent, properties of the systems of circuits $\mathscr{C}(Y, X)$, where X and Y are both independent, and on this basis to analyze how 'far' an independence system is from being a matroid. In fact, it can be shown (see [3]) that in the case of $p = 1$ and $r = 1$ conditions (i) and (ii) are equivalent and, therefore, matroids are just (1, 1)-systems. Thus, (p, r)-systems provide a classification scheme for independence systems, in particular for those, which are closely related to well known combinatorial optimization problems such as travelling salesman problems, vertex-packing in finite graphs or the matchoid problem. For a related investigation as well as other axiomatic foundations of (p, r)-systems the reader is referred to [3].

Furthermore, for the case of $p = 2$ and $r = 2$, conditions (i) resp. (ii) reflect those properties which play an important role for the existence of augmenting paths when determining a maximum vertex-packing in a claw-free graph or a maximum independent set in two matroids. We will come back to this aspect later.

The basic subject of this paper is the class of (2, 2)-systems. Within the hierarchy of (p, r)-systems, given by the positive integers p and r, they directly follow matroids, since $(t, 1)$- or $(1, t)$-systems do not exist for $t > 1$: by the equivalence of conditions (i) and (ii) for $p = 1$ and $r = 1$ and by the minimality of p and r, $p = 1$ resp. $r = 1$ would imply $r = 1$ resp. $p = 1$. However, as the following example shows, there exist (p, r)-systems, where p and r are not equal.

Example 2.2. Let $(E, \mathscr{I})$ be given by $E = \{1, 2, 3, 4, 5\}$ and by its system of circuits $\{\{1, 2, 4\}, \{2, 3\}, \{3, 4\}, \{2, 5\}, \{4, 5\}\}$. It is easy to check that $(E, \mathscr{I})$ is a (p, r)-system with $p = 2$, but $X = \{2, 4\}$, $Y = \{1, 3, 5\}$ yield $|\mathscr{C}(Y, X)| = 5 > 2|\{2, 4\}| = 4$; hence $r > p$.

To make the reader more acquainted with the concept of (2, 2)-systems, we will first review some examples, which have been studied more or less intensively within the last few years. A detailed investigation can be found in [3].

2.1. 2-Matroid intersection

Let $M(E, \mathcal{I}_1)$, $M(E, \mathcal{I}_2)$ be two matroids. Then the independence system $(E, \mathcal{I})$, where $\mathcal{I}$ is given by $\mathcal{I}_1 \cap \mathcal{I}_2$ is called a *2-matroid intersection*. It could happen that $(E, \mathcal{I})$ is again a matroid, i.e., a (1, 1)-system. In general, however, $(E, \mathcal{I})$ is a (2, 2)-system.

The problem of finding a maximum independent set in a 2-matroid intersection in polynomial time has been solved by Edmonds [1] and Lawler [6]. We note that the latter approach is based upon a concept of augmenting paths in a special digraph, which goes back to Krogdahl (cf. Lawler [6]).

2.2. Vertex-packing in claw-free graphs

Let $G = (E, \mathscr{C})$ be a finite, simple graph with vertex set E and edge set $\mathscr{C}$. The independence system $(E, \mathcal{I})$ consisting of all subsets of pairwise non-adjacent vertices is called its *vertex-packing independence system*, which, for the sake of brevity, will be called VP-independence system. If G does not contain the complete, bipartite graph $K_{1,3}$ as induced subgraph, then G is said to be *claw-free*. In particular, its VP-independence system is a matroid, iff G does not contain an induced $K_{1,2}$. Otherwise it is a (2, 2)-system.

Based upon a concept of augmenting paths polynomial algorithms for finding a maximum vertex-packing in a claw-free graph G have been designed by Minty [8] and Sbihi [10].

2.3. Claw-free graph—matroid intersection

Again we consider a finite, simple, claw-free graph $G = (E, \mathscr{C})$. For any vertex $e \in E$ we define

$$N(e) := \{e' \in E : \exists C \in \mathscr{C} \;\; C = \{e, e'\}\},$$

the *neighbourhood* of e, and we look at

$$E' := \{e \in E : N(e) \text{ is a clique}\}.$$

Now we define an arbitrary matroid $M(E', \mathcal{I}')$ over E' and extend it to the whole ground set E by adjoining all $e \in E \backslash E'$ as free elements, which yields a matroid $M(E, \mathcal{I}_M)$. It can be shown that $(E, \mathcal{I})$, $\mathcal{I}$ given by $\mathcal{I} := \mathcal{I}_M \cap \mathcal{I}_G$, where $(E, \mathcal{I}_G)$ is the VP-independence system of G, is either again a matroid or a (2, 2)-system.

A polynomial algorithm to solve the related maximum independent set problem is not known. However, claw-free graph–matroid intersections generalize the concept of matching-forests as introduced by Giles [4]. For let $G^1 = (V, E)$ be a finite, loopless, 'mixed' graph, i.e., let E consist of directed as well as undirected edges. Then a *matching-forest* in G^1 is a forest in G^1 such that for every vertex $v \in V$ the number of edges directed to v plus the number of undirected edges incident to v is at most 1. It is easily shown that for X being a matching-forest it is sufficient that the directed edges of X constitute a forest and that the condition on the incidence at the vertices as stated above holds. Now consider the set E^1 of all those directed edges $e = (u, v)$ such that there exists an undirected (or directed, but different) edge e^1, which is incident to (directed to or from) the vertex u. For any such edge we add a new vertex u_e to V and replace $e = (u, v)$ by $e = (u_e, v)$.

Finally we delete the directions of all directed edges. We obtain an undirected graph G^2 and observe that the matching-forest independence system of G^1 is now given by the intersection of the matching independence system of G^2 and the graphic matroid over that subgraph of G^1, which is induced by the directed edges of G^1.

Note, that for any of the corresponding undirected edges in G^2 there is always one endpoint to which only this edge is incident. If we now go over to the line-graph G of G^2 we can represent the matching-forest independence system of G^1 by the intersection of a matroid and the VP-independence system of a claw-free graph exactly as introduced above.

A polynomial algorithm to find a maximum matching-forest in a mixed graph has been designed in [4], but a straightforward generalization of augmenting paths is not its basis.

2.4. *b*-Matchings

Here we consider a finite, loopless graph $G = (V, E)$ (multiple edges are allowed) together with a mapping $b \in N^V$, which assigns to each vertex $v \in V$ a positive integer b_v, the capacity of vertex v. A set of edges $X \subseteq E$ is called a *b-matching* or degree-constrained subgraph of G, if for all $v \in V$ the number of edges of X incident to v is less or equal to b_v. Obviously, the system of all b-matchings of G relative to the mapping b constitutes an independence system $(E, \mathcal{I})$, which, once again, is either a matroid (a rather special case) or a (2, 2)-system.

The problem of finding a maximum b-matching in G in polynomial time has been solved by Edmonds [2].

2.5. Matchoids

As before let $G=(V,E)$ be a finite, loopless graph possibly containing multiple edges. For any vertex $v \in V$ and any subset A of E we define

$$A(v) := \{e \in A: e \text{ is incident to } v\}.$$

Now, for each vertex $v \in V$ let there be given a matroid on $E(v)$, say $M(E(v), \mathcal{I}(v))$. The independence system $(E, \mathcal{I})$, obtained via

$$\mathcal{I} := \{X \subseteq E: X(v) \in \mathcal{I}(v)\ \forall v \in V\}$$

is called a *matchoid* on G. We point to the fact that the enlargement of $M(E(v), \mathcal{I}(v))$ to $M(E, \mathcal{I}_v)$ by adjoining the elements $e \in E \backslash E(v)$ as free elements provides a common ground set for all matroids and thus $\mathcal{I} = \bigcap_{v \in V} \mathcal{I}_v$.

When all matroids $M(E(v), \mathcal{I}(v))$ are b_v-uniform, $(E, \mathcal{I})$ becomes a b-matching independence system, and if G consists of 2 vertices together with the multiple edges E we obtain the case of a 2-matroid intersection. Matchoids are either $(1,1)$-systems or $(2,2)$-systems.

Although they have been studied by several authors (cf. [5, 6, 7]), a polynomial algorithm to find a maximum independent set in an arbitrary matchoid has not yet been found. Up to now such an algorithm has been developed for the case of polymatroids, which are represented as 2-dimensional subspaces of a linear space, by Lovàsz [7]. The purpose of this study is to give some idea how to carry over the concept of augmenting paths to the general case of $(2,2)$-systems, thus suggesting an approach to the problem of determining a maximum independent set in a $(2,2)$-system in polynomial time.

3. The dependence graph

The concept of the dependence graph has been introduced within the framework of 2-matroid intersection algorithms by Krogdahl and Lawler [6]. To allow a more general point of view we will modify their concept slightly.

Definition 3.1. Let $(E, \mathcal{I})$ be a $(2,2)$-system and $X \in \mathcal{I}$. Then for X and $Z \subseteq E \backslash X$ the *dependence graph* DPG(Z, X) is a bipartite graph with vertex set $V = Z \cup X$ and edge set K as follows:

(i) $\{e, e'\} \in K$ if and only if $e \in Z$ and $e' \in C$ for some circuit $C \in \mathscr{C}(e, X)$;

(ii) an edge $\{e, e'\}$ is doubled if and only if there exist $C_1, C_2 \in \mathscr{C}(e, X)$ such that $C_1 \neq C_2$ and $\{e, e'\} \subseteq C_1 \cap C_2$; one of the edges is then assigned to C_1, the other to C_2.

For algorithmic reasons it could be useful to color the edges corresponding to the possibly two different circuits in $X \cup \{e\}$ by two different colors. We note that every circuit $C \in \mathscr{C}(Z, X)$ is 'taken up' once in $\mathrm{DPG}(Z, X)$ as described.

The construction of $\mathrm{DPG}(Z, X)$ can be performed in polynomial time (depending on testing for independence) for any $X \in \mathscr{I}$ and $Z \subseteq E \backslash X$:

(i) If $X \cup \{e\} \in \mathscr{I}$, nothing has to be done.

(ii) If $X \cup \{e\} \notin \mathscr{I}$, delete from $X \cup \{e\}$ elements e' as long as dependence is guaranteed. The resulting subset of $X \cup \{e\}$ is a circuit C_1.

(iii) If there exists $e' \in C_1$ such that $X \cup \{e\} \backslash \{e'\} \notin \mathscr{I}$, then for $X \cup \{e\} \backslash \{e'\}$ we get a circuit $C_2 \subseteq X \cup \{e\}$ different from C_1 by the same procedure as in (ii).

Definition 3.2. A *W-path* W in $\mathrm{DPG}(Z, X)$, $X \in \mathscr{I}$, $Z \subseteq E \backslash X$, is a path $(e_1, e_2, \ldots, e_{2k+1})$ in $\mathrm{DPG}(Z, X)$ without repetition of vertices such that the following conditions are fulfilled:

(i) $X \cup \{e_i\}$ contains exactly one circuit for $i = 1$ and $i = 2k + 1$;

(ii) those edges of W, which connect e_i and e_{i+1} resp. e_{i+1} and e_{i+2} correspond to different circuits C_1, C_2 both from $\mathscr{C}(e_{i+1}, X)$ (i.e., in case of a coloration of the edges of $\mathrm{DPG}(Z, X)$ these two edges have different colors) for $i = 2, 4, \ldots, 2k - 2$;

(iii) $\{e_1, e_3, \ldots, e_{2k+1}\} \in \mathscr{I}$.

The number $2k + 1 \geqslant 3$ is called the length of W; a single vertex $e \in Z$ such that $X \cup \{e\} \in \mathscr{I}$ is considered as a W-path of length 1.

We note that this definition of a W-path generalizes the notion of an augmenting path in claw-free graphs as well as a source-to-sink path in the bordergraph as considered in [6] for 2-matroid-intersections; the latter in the case that any $C \in \mathscr{C}(W \backslash X, X)$ is either a circuit of the matroid M_1 or of M_2.

Associated with a W-path W of length $2k + 1$ are $2k$ different circuits, every one of which is *represented* by an edge in W.

Theorem 3.3. *Let* $(E, \mathscr{I})$ *be a* $(2,2)$*-system and* $X, Y \in \mathscr{I}$, $|X| + 1 = |Y|$. *Then there is a* W*-path* W *in* $\mathrm{DPG}(Y \backslash X, X)$.

Proof. If there exists an element $e \in Y \backslash X$ such that $X \cup \{e\} \in \mathscr{I}$, $W = \{e\}$, a W-path of length 1. So suppose that $\mathscr{C}(e, X) \neq \emptyset$ for all $e \in Y \backslash X$.

By definition of a $(2,2)$-system, the family $(C \backslash Y, C \in \mathscr{C}(Y, X))$ can be partitioned into 2 nonempty subfamilies $(C \backslash Y, C \in \mathscr{C}_1(Y, X))$ and $(C \backslash Y, C \in \mathscr{C}_2(Y, X))$, each of which possesses a transversal, say T_1 and T_2, respectively. In particular, there exist two bijections $\pi_1 : T_1 \to \mathscr{C}_1(Y, X)$, $\pi_2 : T_2 \to \mathscr{C}_2(Y, X)$ such that $x \in \pi_i(x) \backslash Y$ for all $x \in T_i$ and $i = 1, 2$, respectively.

Now we label for all $e' \in T_i$ that edge between e and e', which represents the circuit $\pi_i(e')$ and where e is given by $\{e\} = \pi_i(e') \cap (Y \backslash X)$, $i = 1, 2$. Consider that subgraph $\bar{G}$ of DPG$(Y \backslash X, X)$, which consists of the vertices $\bar{V} = (X \backslash Y) \cup (Y \backslash X)$ and the labelled edges. We observe that each vertex in $\bar{G}$ has degree $\leqslant 2$. Hence the connected components of $\bar{G}$ consist of isolated vertices, simple paths or simple cycles. Since $|Y \backslash X| = |X \backslash Y| + 1$, one of these components must be a W-path. □

Example 3.4. Let a claw-free graph–matroid intersection $(E, \mathcal{I})$ be given by graph G and the graphic matroid arising from graph G' as shown in Fig. 1. Then $X = \{2, b, c, 4, d, e\}$ and $Y = \{1, f, 3, g, 5, h, 7\}$ are both independent. For DPG(Y, X) we obtain the graph shown in Fig. 2, where the dotted edges represent the matroid circuits. A W-path W in DPG(Y, X) is $W = (1, 2, f, b, 3, c, g, 4, 5)$.

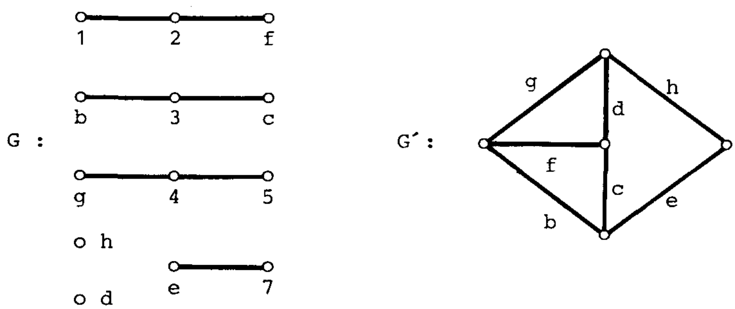


Fig. 1.

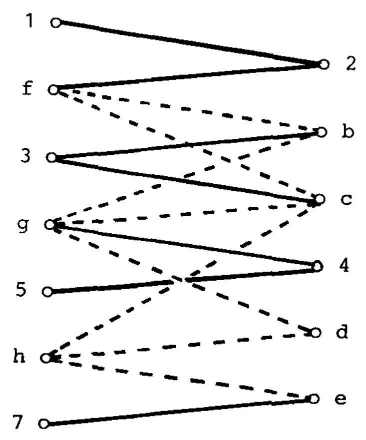


Fig. 2.

Corollary 3.5 (Sufficient Optimality Criterion). *Let $(E, \mathcal{I})$ be a (2,2)-system and $X \subseteq E$ be a base of E such that there exists at most one element $e \in E \backslash X$ with the property*

$X \cup \{e\}$ *contains exactly one circuit.*

Then X is of maximal cardinality within $\mathcal{I}$.

Corollary 3.5 as well as Theorem 3.3 in general do not hold for other than (1,1)- and (2,2)-systems. The example illustrated in Fig. 3 goes back to Sakarovitch (cf. [10]). The edges of this hypergraph are considered to be the circuits of an independence system $(E, \mathcal{I})$. One easily verifies that for any $X \in \mathcal{I}$, $e \in E \backslash X$, $|\mathcal{C}(e, X)| \leq 2$. In particular for $X = \{3, 4, 6, 7\}$ and any $e \in Y = \{1, 2, 5, 8, 9\}$, $|\mathcal{C}(e, X)| = 2$. However, $(C \backslash Y, C \in \mathcal{C}(Y, X)) = (\{3\}, \{4\}, \{3\}, \{4\}, \{3, 4\}, \{6, 7\}, \{6\}, \{7\}, \{6\}, \{7\})$ and $|\mathcal{C}(Y, X)| = 10 > 2|\{3, 4, 6, 7\}|$. Thus $(E, \mathcal{I})$ is not a (2,2)-system and no W-path does exist in $\mathrm{DPG}(Y \backslash X, X)$.

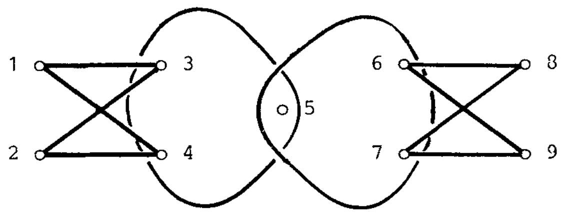


Fig. 3.

4. The exchange $(X \backslash W) \cup (W \backslash X)$

As the algorithms for 2-matroid intersection and claw-free graphs show, trying to find a W-path in $\mathrm{DPG}(E \backslash X, X)$ for some known X seems to be one of the essential features. Before looking at the general case of arbitrary (2,2)-systems let us review the situation for the examples presented above and derive some further results.

4.1. 2-Matroid intersection

For any $X, Y \in \mathcal{I}$ we can partition the family $(C \backslash Y, C \in \mathcal{C}(Y, X))$ into two subfamilies $(C \backslash Y, C \in \mathcal{C}_1(Y, X))$, $(C \backslash Y, C \in \mathcal{C}_2(Y, X))$, where $\mathcal{C}_1(Y, X)$, $\mathcal{C}_2(Y, X)$ are circuits of the matroids $M(E, \mathcal{I}_1)$, $M(E, \mathcal{I}_2)$. Clearly, both families possess a transversal, say T_1 and T_2. Now we can represent $\mathcal{C}_1(Y, X)$ resp. $\mathcal{C}_2(Y, X)$ by a matching M_1 resp. M_2 in $\mathrm{DPG}(Y \backslash X, X)$ and the union of these two matchings represents the system $\mathscr{C}(Y, X)$. Since $|Y| = |X| + 1$, there exists a W-path W. If C is a circuit in both matroids and $C \in \mathscr{C}(E\backslash X, X)$, then one may add the corresponding edges in DPG($E\backslash X, X$) twice, which guarantees that edges in W, which represent a circuit from $M(E, \mathscr{I}_1)$ or $M(E, \mathscr{I}_2)$ follow each other alternately. This allows to shortcut W, as described in [6] to obtain a W-path W', for which $(X\backslash W') \cup (W'\backslash X) \in \mathscr{I}$ holds.

In general (2, 2)-systems shortcutting is not as efficient. We do not necessarily obtain a new W-path W' 'shorter' than W, so that the existence of a W-path W fulfilling $(X\backslash W) \cup (W\backslash X) \in \mathscr{I}$ cannot be deduced.

4.2. *Vertex-packing in claw-free graphs*

Due to the special structure of the circuits $\mathscr{C}$ of $(E, \mathscr{I})$, the VP-independence system of a claw-free graph G, the edges in DPG($E\backslash X, X$) correspond exactly to the circuits $C \in \mathscr{C}(E\backslash X, X)$. Thus a W-path W in DPG($E\backslash X, X$) allows the augmentation from X to $\bar{X} := (X\backslash W) \cup (W\backslash X)$ and $|\bar{X}| = |X| + 1$.

4.3. *Claw-free graph—matroid intersection*

This is one of those cases, where the possibility of shortcutting a W-path W is limited. If we restrict ourselves to W-paths, where those edges, which represent the circuits of the matroid, form a matching (this is always possible), then the following may occur:

> In W there is an even number of edges of G between two consecutive edges representing matroid circuits.

Then main cycles, as introduced in [6], cannot be eliminated by shortcutting. Moreover, $\bar{X} := (X\backslash W) \cup (W\backslash X)$ can contain one or even more distinct circuits of the matroid.

4.4. *b-Matchings*

For graphs $G = (V, E)$ possessing no multiple edges we can derive a result, which could be used to generalize the concept of augmenting paths as known from 1-matchings to a concept of 'augmenting subgraphs'. First, we reformulate Theorem 3.3 for this type of b-matching independence systems to obtain a W-path possessing some specific structure.

Theorem 4.1. *Let $(E, \mathscr{I})$ be a b-matching independence system for the simple, finite graph $G = (V, E)$ and $X, Y \in \mathscr{I}$ such that $|Y| = |X| + 1$. Then there is a W-path $W = (e_1, \ldots, e_{2k+1})$ in* DPG($Y\backslash X, X$) *with the following property:*

(P) *Those edges of W, which represent the system $\mathscr{C}_v(W\backslash X, X) \subseteq \mathscr{C}_v$, the system of circuits arising at vertex v, form a matching M_v for all $v \in V$.*

Proof. If $W = \{e\}$, nothing has to be shown. So let $2k+1 > 1$.

If $X \cup \{e\}$, $e \in Y\backslash X$, contains a circuit $C \in \mathscr{C}_v$, then e must be incident to vertex v and $|X(v)| = b_v$. Since G is simple and $b_v > 0$ for all $v \in V$, there exists for each circuit $C \in \mathscr{C}(Y, X)$ a unique vertex v such that $C \in \mathscr{C}_v$. Hence we can partition $\mathscr{C}(Y, X)$ into at most $2|X|$ many systems $\mathscr{C}_v(Y, X)$ each of them being a subsystem of $\mathscr{C}_v$.

Moreover, each family $(C\backslash Y, C \in \mathscr{C}_v(Y, X))$ possesses a transversal T_v and the labelling of the edges $\{e, e'\}$, where $e' \in T_v$ and e is given by $\pi_v(e') \cap (Y\backslash X)$, yields a matching M_v in DPG$(Y\backslash X, X)$, which represents the system $\mathscr{C}_v(Y, X)$.

Since every edge is incident to at most 2 distinct vertices in DPG$(Y\backslash X, X)$, $\bigcup_{v \in V} M_v$ induces a subgraph $\bar{G}$, one of whose components must be a W-path $W = (e_1, \ldots, e_{2k+1})$, which by construction of $\bar{G}$ fulfills property (P). $\square$

Let us look now at the exchange $(X\backslash W) \cup (W\backslash X)$.

Theorem 4.2. *Let $(E, \mathscr{I})$ be as in Theorem 4.1 and W a W-path in* DPG$(Y\backslash X, X)$, *where $X, Y \in \mathscr{I}$, $|X| + 1 = |Y|$; W is assumed to fulfill* (P). *Then $(X\backslash W) \cup (W\backslash X)$ contains at most one circuit.*

Proof. Let M_v be the matching, which represents the system $\mathscr{C}_v(W\backslash X, X) \neq \emptyset$ for some $v \in V$. Then for any other $e \in W$, which is not covered by an edge of M_v, either two edges from matchings $M_u \neq M_v$, $M_w \neq M_v$, $u \neq w$, are incident to e, or one edge from a matching M_u, $u \neq v$, is incident to e.

Candidates for the latter situation are e_1 or e_{2k+1}. For the vertices w_1 resp. w_{2k+1}, both $\neq u$, which e_1 resp. e_{2k+1} is incident to in G, the conditions

$$|X \cup \{e_1\}(w_1)| \leq b_{w_1}, \qquad |X \cup \{e_{2k+1}\}(w_{2k+1})| \leq b_{w_{2k+1}}$$

hold.

Thus all $e \in W$, which are not covered by M_v, are not incident to vertex v in G. Hence,

$$|(X\backslash W) \cup (W\backslash X)(v)| = b_v \quad \text{for all vertices } v \in V \text{ with } \mathscr{C}_v(Y, X) \neq \emptyset.$$

However, if e_1 and e_{2k+1} are both incident to a vertex $\bar{v}$ in G, i.e., $w_1 = w_{2k+1} = \bar{v}$ and if $|X(\bar{v})| = b_{\bar{v}} - 1$, then $|(X\backslash W) \cup (W\backslash X)(\bar{v})| = b_{\bar{v}} + 1$, which means that $(X\backslash W) \cup (W\backslash X)$ contains a circuit $C \in \mathscr{C}_{\bar{v}}$, which is unique.

If $w_1 \neq w_{2k+1}$ or if $|X(\bar{v})| \leq b_{\bar{v}} - 2$, then $(X\backslash W) \cup (W\backslash X) \in \mathscr{I}$. $\square$

Example 4.3. Let a b-matching independence system $(E, \mathscr{I})$ be given by graph G in Fig. 4 together with the mapping $b: V \to \{1, 2\}$, which assigns the number 1 to vertices $v_1, v_2, v_3, v_5, v_6, v_7$ and the number 2 to v_4. Then for $X = \{2, 5, 6\}$, $Y = \{1, 3, 4, 7\}$, both independent sets, we obtain as $\mathrm{DPG}(Y\backslash X, X)$ the second graph of Fig. 4. We observe that $W = (3, 2, 1, 5, 4)$ is a W-path in $\mathrm{DPG}(Y\backslash X, X)$, but the exchange $(X\backslash W) \cup (W\backslash X) = \{1, 3, 4, 6\}$ yields a dependent set. However, the exchange $(X\backslash W') \cup (W'\backslash X)$ for $W' = \{3, 2, 1, 5\}$ yields the independent set $\{1, 3, 6\}$. For this set and Y we obtain another path $W'' = (4, 6, 7)$ and the corresponding exchange yields Y. Thus we have augmented by exchanging twice a W-path resp. parts of a W-path.

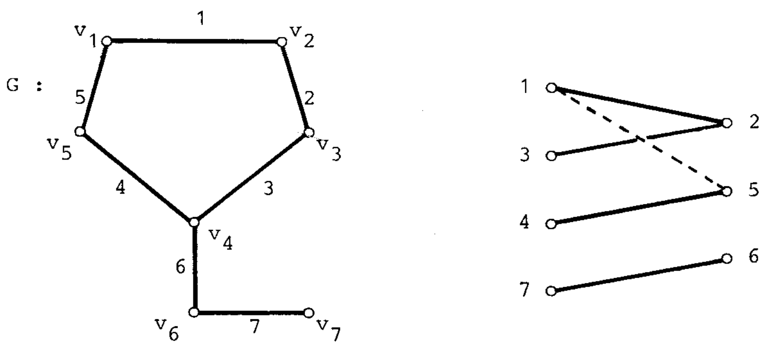


Fig. 4.

4.5. Matchoids

By an almost identical argumentation as used in Theorem 4.1 we can show the following theorem.

Theorem 4.4. *Let $(E, \mathscr{I})$ be a matchoid on $G = (V, E)$, a simple, finite graph, and $X, Y \in \mathscr{I}$ such that $|Y| = |X| + 1$. Then there is a W-path $W = (e_1, \ldots, e_{2k+1})$ in $\mathrm{DPG}(Y\backslash X, X)$ such that*:

(Q) *Those edges of W, which represent the system $\mathscr{C}_v(W\backslash X, X)$, form a matching for all $v \in V$.*

However, as the following example shows, $(X\backslash W) \cup (W\backslash X)$ can contain more than one circuit.

Example 4.5. Let the matchoid $(E, \mathscr{I})$ be given by graph $G = (V, E)$ of Fig. 5, where $V = \{v_1, \ldots, v_{12}\}$ and $E = \{1, \ldots, 13\}$, together with a number of k-uniform matroids as induced by the mapping $b: V\backslash\{v_5\} \to \{1, 2\}$, which assigns

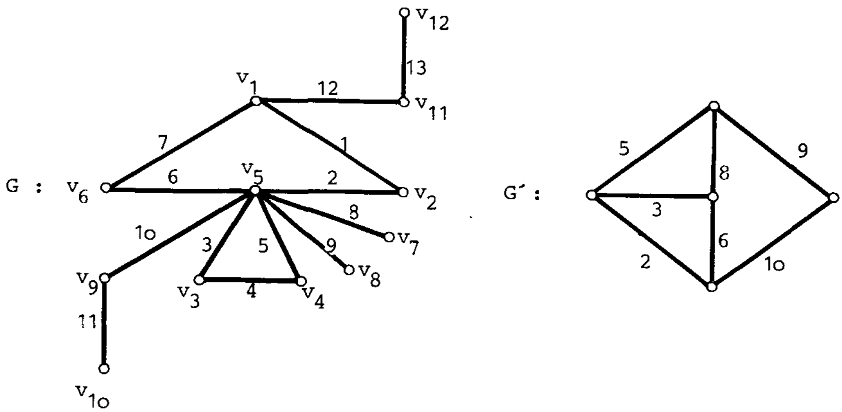


Fig. 5.

the number 1 to each vertex v_i, $i \in \{1, \ldots, 12\} \backslash \{1, 5\}$ and the number 2 to v_1. Furthermore, let a graphic matroid $M(E(v_5), \mathscr{I}(v_5))$ be defined by graph G' of Fig. 5. $X = \{2, 4, 6, 8, 10, 12\}$ and $Y = \{1, 3, 5, 7, 9, 11, 13\}$ are both independent in the matchoid $(E, \mathscr{I})$. $W = (1, 2, 3, 4, 5, 6, 7)$ is a W-path in DPG(Y, X) fulfilling (Q), but $(X \backslash W) \cup (W \backslash X) = \{1, 3, 5, 7, 8, 10, 12\}$ contains the two circuits $\{3, 5, 8\}$ and $\{1, 7, 12\}$.

5. W-paths and augmentation

The question arises how to use a single or several W-paths for the determination of a maximum independent set in a (2, 2)-system, resp. the examples presented in Sections 2.3, 2.4 and 2.5.

First, let $(E, \mathscr{I})$ be a (2, 2)-system and X, Y be two independent sets such that $|Y| = |X| + 1$. From Theorem 3.3 we know that there is a W-path W^1 in DPG$(Y \backslash X, X)$, whose set of vertices is contained in $(X \backslash Y) \cup (Y \backslash X)$. If $(X \backslash W^1) \cup (W^1 \backslash X)$ is independent, we can augment. Otherwise we can enlarge $X \backslash W^1$ by a subset $\bar{W}^1$ of $W^1 \backslash X$ to a base X^2 of $(X \backslash W^1) \cup (W^1 \backslash X)$. We observe

- $\bar{W}^1$ is nonempty;
- $|X^2| \leq |X| < |Y|$;
- $|X^2 \cap Y| > |X \cap Y|$.

Consequently, in DPG$(Y \backslash X^2, X^2) =$ DPG$(Y \backslash (X \cup \bar{W}^1), X^2)$ there exists a W-path $W^2 \subseteq (Y \backslash X^2) \cup (X^2 \backslash Y)$. In particular, $W^2 \cap \bar{W}^1 = \emptyset$. Again, we can check, whether $(X^2 \backslash W^2) \cup (W^2 \backslash X^2)$ is independent and, if this is not the case, go over to $(X^2 \backslash W^2) \cup \bar{W}^2$ until we have reached a number m such that the condition

$$\sum_{i=1}^{m} |\bar{W}^i| = \left(\sum_{i=1}^{m} |X \cap W^i|\right) + 1$$

is fulfilled. Then we have augmented. Since

$$|X^i \cap Y| < |X^{i+1} \cap Y| \quad \text{for all } 1 \leqslant i \leqslant m,$$

where $X^1 = X$, m is bounded from above by $|Y\backslash X|$ ($\leqslant \lfloor\frac{1}{2}|E|\rfloor + 1$).

Now let us specify how the knowledge of Y can be avoided for an augmentation of X. We have to determine sequences $W^1, W^2, \ldots, W^m$ and $\bar{W}^1, \bar{W}^2, \ldots, \bar{W}^m$ such that for $i = 1, \ldots, m$

- W^i is a W-path in $\mathrm{DPG}(E\backslash[X \cup (\bigcup_{j=1}^{i-1} \bar{W}^j)], X^i)$ and $W^i \cap (\bigcup_{j=1}^{i-1} \bar{W}^j) = \emptyset$,
- $\bar{W}^i$ is a subset of $W^i\backslash X^i$ such that $X^{i+1} := (X^i\backslash W^i) \cup \bar{W}^i$ is a base of $(X^i\backslash W^i) \cup (W^i\backslash X^i)$,

and

$$\sum_{j=1}^{i} |\bar{W}^j| \leqslant \sum_{j=1}^{i} |X \cap W^j| \quad \text{for } i = 1, \ldots, m-1,$$

but

$$\sum_{j=1}^{m} |\bar{W}^j| = \left(\sum_{j=1}^{m} |X \cap W^j|\right) + 1.$$

The essential part of an algorithm for the determination of a maximum independent set in a (2, 2)-system would consist in determining such sequences relative to an independent set X.

6. Conclusions

We have generalized the concept of an augmenting path from special examples to the general case of (2, 2)-systems and indicated, how an independent set X could be augmented by using a finite sequence of such generalized augmenting paths, which have been called W-paths. We could show in Theorem 3.3 that in case of $X, Y \in \mathcal{J}$, $|Y| = |X| + 1$, a W-path always exists in $\mathrm{DPG}(Y\backslash X, X)$ and thus in $\mathrm{DPG}(E\backslash X, X)$. Since the number of such W-paths W^i needed for augmentation is bounded from above by $\lfloor\frac{1}{2}|E|\rfloor + 1$ and due to the special restrictions on W^i for $i = 2, \ldots, m$ we believe that there is a polynomial algorithm for the determination of a maximum independent set in a (2, 2)-system.

As far as the design of such an algorithm for the special cases of a matchoid or the intersection of a matroid with the VP-independence system of a claw-free graph is concerned it would be interesting to investigate the corresponding sequences of W-paths with respect to their special structure.

Acknowledgement

I would like to thank the referees for their valuable suggestions.

References

[1] J. Edmonds, Matroid-intersection, Ann. of Discrete Math. 4 (1979) 39–49.
[2] J. Edmonds, An introduction to matching, Mimeographed Lecture Notes, Engrg. Summer Conf. Ann Arbor, Univ. of Michigan, 1967.
[3] R. Euler, On a classification of independence systems, Rept. 7/80, Mathematisches Institut der Universität zu Köln, 1980.
[4] R. Giles, Optimum matching forests I, Dept. of Math., Univ. of Kentucky, 1977.
[5] T.A. Jenkyns, Matchoids: A generalization of matchings and matroids, Thesis, Univ. of Waterloo, 1974.
[6] E.L. Lawler, Combinatorial Optimization: Networks and Matroids (Holt, Rinehart and Winston, New York, 1976).
[7] L. Lovàsz, The matroid parity problem, Univ. of Waterloo, 1979.
[8] G.J. Minty, On maximal independent sets of vertices in claw-free graphs, J. Combin. Theory Ser. B 28 (1980) 284–304.
[9] L. Mirsky, Transversal Theory (Academic Press, New York, 1971).
[10] N. Sbihi, Etude des stables dans les graphes sans etoile, Thèse, Univ. Scientifique et Médicale de Grenoble, 1978.

Annals of Discrete Mathematics 16 (1982) 83–95

TRANSFORMATIONS WHICH PRESERVE PERFECTNESS AND H-PERFECTNESS OF GRAPHS

J. FONLUPT and J.P. UHRY

I.M.A.G., B.P. 53 X, 38041 Grenoble Cedex, France

A graph G is h-perfect if the convex hull of the incidence vectors of the independent sets of G is a polytope defined by nonnegativity constraints, clique constraints and odd holes constraints.

We prove the two following theorems:

(1) A graph obtained by identification of two vertices of a bipartite graph is h-perfect.

(2) If in a perfect graph there exists two vertices b and c such that all the minimal chains between b and c have an odd number of vertices, the graph obtained by identification of b and c is perfect.

1. Introduction

Let $G = (V, E)$ be a finite, undirected, loopless graph; V is the set of vertices of G and E denotes the set of edges of G. We define $P(G)$ as the convex hull of the incidence vectors of independent sets of G. $P(G)$ is a polytope and therefore can also be characterized by a set of inequality constraints.

It is well known that a graph G is perfect [4] if and only if $P(G)$ is defined by the following constraints:

(1) nonnegativity constraints on the variables associated with the vertices of G.

(2) clique constraints.

When G has odd holes, it is also necessary to introduce a third type of constraint:

(3) odd hole constraints.

Of course, constraints (1), (2) and (3) are in general not sufficient to characterize $P(G)$ for a given graph G. If this characterization is sufficient, we shall call the graph G an h-perfect graph. Chvatal [3] was the first author to be interested in h-perfect graphs and some interesting classes of h-perfect graphs have been studied in [2] and [5].

In this paper, we study simple operations on graphs which preserve perfection (Section 3) and h-perfection (Section 4).

In Section 2, we give some definitions and define some notation we shall use throughout this study.

2. Definitions and notation

(a) A *chain* from $a \in V$ to $b \in V$ is a sequence $T = (v_1, \ldots, v_k)$ of not necessarily distinct vertices such that:
- $v_1 = a$;
- $v_k = b$;
- $(v_i, v_{i+1}) \in E \ \forall\, 1 \leq i \leq k$.

A chain T from $a \in V$ to $b \in V$ is *minimal* if there exists no proper subsequence of T which is a chain from a to b.

A *pseudo-cycle* of G is a sequence of vertices $D = (v_1, \ldots, v_k)$ such that:
- D is a chain from v_1 to v_k:
- $(v_k, v_1) \in E$.

(v, w) is a *chord* of D if v and w are adjacent vertices of G but not consecutive vertices of D.

A pseudo-cycle is odd if its *length* $|D|$ is odd.

$D(G)$ will denote the set of odd pseudo-cycles of G.

A chordless pseudo-cycle with distinct vertices is a *hole.*

$C(G)$ will denote the set of odd holes of G.

Finally, let us call $S(G)$ the family of *independent sets* of G and $\Omega(G)$ the family of *maximal cliques* of G.

(b) x represents a vector of $R^{|V|}$ and to each vertex $v \in V$ we associate one coordinate of x noted x_v.

The incidence vector x^S of a subset S of V is defined by
- $x_v^S = 1$ if $v \in S$.
- $x_v^S = 0$ if $v \notin S$.

If $T = (v_1, \ldots, v_k)$ is a sequence of not necessarily distinct vertices of V, we shall set

$$x(T) = \sum_{i=1}^{k} x_{v_i}\,.$$

Note that if $v \in V$, x_v will appear in the description of $x(T)$ as many times as v appears in the description of T.

For example, the pseudo-cycle $D = (a, b, c, d, e, f, c, g)$ described by Fig. 1 is a pseudo-cycle of length 8 and

$$x(D) = x_a + x_b + 2x_c + x_d + x_e + x_f + x_g\,.$$

(c) A graph G is *perfect* (see [3]) if and only if $P(G)$ is defined by

$$\begin{cases} x \geq 0\,, & (1)\\ \sum_{v \in \Omega} x_v \leq 1 \quad \forall\, \Omega \in \Omega(G)\,. & (2)\end{cases}$$

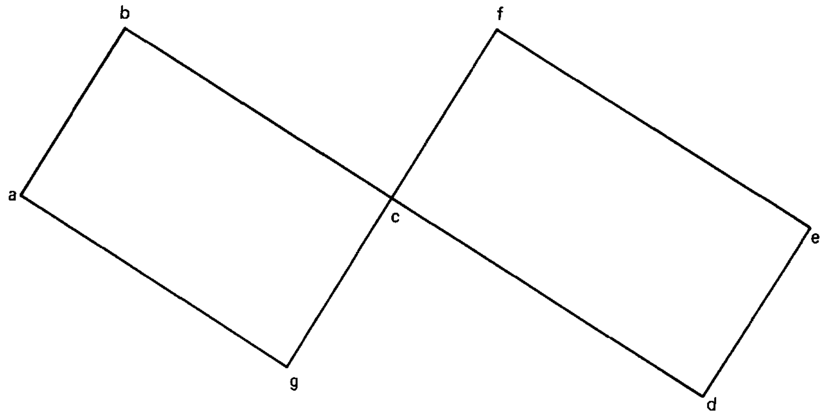


Fig. 1.

A graph G is *h-perfect* if and only if $P(G)$ is defined by

$$\begin{cases} x \geqslant 0\,, & (3)\\ \sum_{v\in\Omega} x_v \leqslant 1 \qquad \forall\, \Omega \in \Omega(G)\,, & (4)\\ x(C) \leqslant \tfrac{1}{2}(|C|-1) \quad \forall\, C \in C(G)\,. & (5)\end{cases}$$

It is obvious that the incidence vector x^S of an independent set of G is an extreme point of the polytope defined by constraints (1) and (2) (resp. (3), (4), (5)). It is also obvious that if x is an integer extreme point of the polytope defined by (1), (2) (resp. (3), (4), (5)), x is the incidence vector of an independent set of G. Therefore, in order to prove that a graph is perfect (resp. h-perfect) we need to prove that all the extreme points of the polytope defined by (1), (2) (resp. (3), (4), (5)) are integer (in fact 0–1) points.

3. A new class of h-perfect graphs

3.1. Definition

Let $G = (V_1, V_2, E)$ be a bipartite graph with b, c two non-adjacent vertices of G. If we delete b and c from G and add a new vertex a whose adjacent vertices are the union of adjacent vertices of b and c, we get a new graph $\bar{G} = (\bar{V}, \bar{E})$. We call $\bar{G}$ the graph obtained from G by identification of b and c.

In this section we want to prove that $\bar{G}$ is a h-perfect graph (Theorem 1 of

Section 3.2). This result is true if $\bar{G}$ is a bipartite graph. Therefore, we shall suppose from now on that $\bar{G}$ is not a bipartite graph. This implies that

(1) One of the two vertices b, c for instance b belongs to V_1 and the other c, belongs to V_2.

(2) There exists a chain $(b, \ldots, c)$ in G.

We remark that, if $T = (b, v_2, \ldots, v_{2k+1}, c)$ is a chain in G between b and c, we can associate to T, after identification of b and c an odd pseudo-cycle $D = (a, v_2, \ldots, v_{2k+1})$ of $\bar{G}$.

Conversely any odd pseudo-cycle of $\bar{G}$ contains a and can be written as $D = (a, v_2, \ldots, v_{2k+1})$ with $v_{2i} \in V_2$, $1 \leqslant i \leqslant k$, and $v_{2i+1} \in V_1$, $1 \leqslant i \leqslant k$.

Therefore, we can associate to D a chain T between b and c in G:

$$T = (b, v_2, \ldots, v_{2k+1}, c).$$

Note that $|D| = |T| - 1$.

The two following lemmas are obvious.

Lemma 1. *Every subgraph of $\bar{G}$ is either a bipartite graph or a graph obtained from a subgraph of G by identification of b and c.*

Lemma 2. *The maximum cardinality of a clique of $\bar{G}$ is less than or equal to three.*

3.2. *The main result*

Theorem 1. *The graphs obtained by identification of two nonadjacent vertices of bipartite graphs are h-perfect.*

Before proving this theorem, we make some useful remarks and prove two preliminary lemmas (Section 3.3). In Section 3.4 we present a labelling procedure for the vertices of G. Then, some results related to this procedure are stated. The final proof of Theorem 1 (Section 3.5) is an easy consequence of these last results. In Section 3.5 we prove a corollary of Theorem 1.

3.3. *Remarks and preliminary lemmas*

Considering cliques of cardinality 3 as holes of length 3, Theorem 1 is equivalent to showing (by Lemma 2), that $P(\bar{G})$ is identical to the polytope

$$P_{\bar{G}} \begin{cases} x_v \geqslant 0 & \forall\, v \in \bar{V}, \\ x_v + x_w \leqslant 1 & \forall\, (v, w) \in \bar{E}, \\ x(C) \leqslant \frac{1}{2}(|C| - 1) & \forall\, C \in C(\bar{G}). \end{cases}$$

More precisely, we just need to prove that all the extreme points of $P_{\bar{G}}$ have 0–1 coordinates.

If Theorem 1 is false, we can assume that, among the graphs which do not satisfy this theorem, $\bar{G}$ has the minimum number of vertices.

Let $x^0 = (x_v^0, v \in V)$ be a non-integer extreme point of $P_{\bar{G}}$.

Lemma 3. $0 < x_v^0 < 1\ \forall\ v \in \bar{V}$.

Proof. First, assume that there exists $w \in \bar{V}$ such that $x_w^0 = 0$. All the feasible solutions of $P_{\bar{G}}$ which satisfy the constraint $x_w = 0$ generate the polytope $P_{\bar{G}_w}$ where $\bar{G}_w$ is obtained from $\bar{G}$ by deletion of w. But by Lemma 1 and our assumption of minimality on $\bar{G}$, $\bar{G}_w$ is an h-perfect graph. This implies that any extreme point of $P_{\bar{G}}$ such that $x_w = 0$ is an integer extreme point. Therefore $0 < x_v^0 \leq 1\ \forall\ v \in \bar{V}$.

From these relations and the clique constraints of $P_{\bar{G}}$, if there exists $w \in \bar{V}$ such that $x_w^0 = 1$, w is an isolated vertex of $\bar{G}$. But this implies that $\bar{G}$ is also an h-perfect graph which is contrary to our assumption.

Lemma 4. *If* $x = (x_v, v \in \bar{V}) \in P_{\bar{G}}$, *then* $x(D) \leq \frac{1}{2}(|D| - 1)\ \forall\ D \in D(\bar{G})$.

Proof. Consider an odd pseudo-cycle which is not a hole of length $2k+1$, $D = (v_1, v_2, \ldots, v_{2k+1})$.

Suppose by induction on k that Lemma 4 is true for all odd pseudo-cycles of length less than $2k+1$.

If a vertex appears more than once in D, we can assume by a suitable indexing of the vertices that

$$D = (v_1, \ldots, v_{2i}, \ldots, v_{2k+1}) \quad \text{with } 1 < i \leq k \text{ and } v_1 = v_{2i}\,.$$

If all the vertices of D are distinct, there exists a chord in D. Again by a suitable indexing of the vertices, D can be written as

$$D = (v_1, \ldots, v_{2i}, \ldots, v_{2k+1}) \quad \text{with } 1 < i \leq k \text{ and } (v_1, v_{2i-1}) \in \bar{E}\,.$$

Thus, if D is not a hole, we can suppose that

$$D_1 = (v_1, \ldots, v_{2i-1}) \in D(\bar{G})\,.$$

As $|D| > |D_1|$,

$$x(D) = x(D_1) + \sum_{j=i}^{k} (x_{v_{2j}} + x_{v_{2j+1}}) \leq \tfrac{1}{2}(|D_1| - 1) + k - i + 1\,.$$

Therefore $x(D) \leq \frac{1}{2}(|D| - 1)$.

3.4. *Labelling procedure on the set of vertices of G*

First, let us set $x_b^0 = x_a^0$ and $x_c^0 = x_a^0$.

Then, we shall define the subset E_1 of saturated edges of E: $(v, w) \in E_1$ if and only if

$$(v, w) \in E \quad \text{and} \quad x_v^0 + x_w^0 = 1.$$

A chain T, from $b \in V_1$ to $v \in V_1 \cup V_2$ is an *alternating chain* if

$$T = (v_1, \ldots, v_k) \quad \text{with } v_1 = b,\ v_k = v$$

and

$$(v_{2i}, v_{2i+1}) \in E_1, \quad 1 < 2i < 2i + 1 \leqslant k.$$

An elementary result we shall use in the next two lemmas is: a chain T from $b \in V_1$ to $v \in V_2$ is an alternating chain if and only if

$$x^0(T) = x_b^0 + x_v^0 + \tfrac{1}{2}(|T| - 2).$$

On Fig. 2, T is an alternating chain of length $2k$. The edges of E_1 are marked by a continuous line.

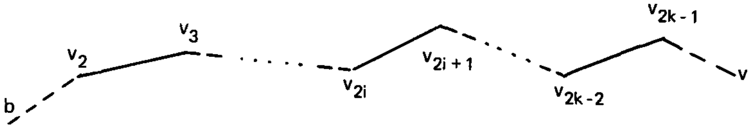


Fig. 2.

A chain from $c \in V_2$ to $v \in V_1 \cup V_2$ is *an alternating chain* if:

$$T = (v_2, \ldots, v_k) \quad \text{with } v_2 = c,\ v_k = v$$

and

$$(v_{2i}, v_{2i+1}) \in E_1, \quad 1 < 2i < 2i + 1 \leqslant k.$$

Starting from a vertex $u \in V_1 \cup V_2$, we can now define a general labelling procedure denoted $P(u)$.

Procedure $P(u)$.

Step 1. Give to u the label $\varepsilon = +1$.

Step 2. If $v \in V_1$ is labelled with ε, label all the still unlabelled vertices adjacent to v with $-\varepsilon$.

Step 3. If $v \in V_2$ is labelled with $-\varepsilon$, label all the still unlabelled vertices w of V_1 such that $(v, w) \in E_1$ with $+\varepsilon$.

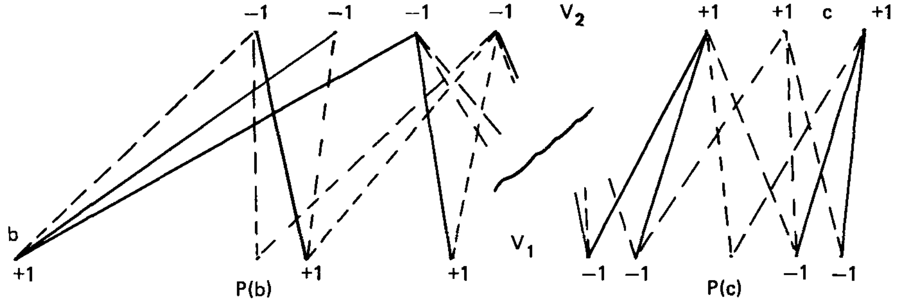


Fig. 3

Repeat Steps 2 and 3 until it is impossible to label more vertices. Note that we shall be concerned with $P(b)$ and $P(c)$. On Fig. 3 edges of E_1 are marked with a continuous line.

Lemma 5. (a) *All the vertices of an alternating chain* $(b, \ldots, v)$ *are labelled by* $P(b)$ *and all the vertices of an alternating chain* $(c, \ldots, v)$ *are labelled by* $P(c)$.

(b) *If* v *is labelled by* $P(b)$ *(resp. by* $P(c)$*), there exists an alternating chain from* b *to* v *(resp. from* c *to* v*).*

Proof. (a) Immediate by the description of Procedure $P(u)$.

(b) Procedure $P(u)$ is similar to an alternating path algorithm for bipartite graphs and this result is classical [1].

Lemma 6. *A vertex* v *cannot be labelled by both* $P(b)$ *and* $P(c)$.

Proof. Suppose that there exists a vertex v which is labelled by $P(b)$ and $P(c)$. Step 2 of Procedure $P(u)$ shows that we can suppose that $v \in V_2$.

By Lemma 5, there exists two alternating chains $(b, \ldots, v)$ and $(c, \ldots, v)$. Concatenating these two chains at v, we obtain a chain T from b to c,

$$T = (v_1, \ldots, v_{2l}, \ldots, v_{2k}) \quad \text{with } v_1 = b,\ v_{2l} = v,\ v_{2k} = c\,.$$

Let D be the odd pseudo-cycle associated with T,

$$x^0(D) = \sum_{i=2}^{2k} x_{v_i}\,.$$

If $1 < l < k$,

$$x^0(D) = \sum_{i=1}^{l-1} (x^0_{v_{2i}} + x^0_{v_{2i+1}}) + x^0_{v_{2l}} + \sum_{i=l}^{k-1} (x^0_{v_{2i+1}} + x^0_{v_{2i+2}})\,.$$

As T_1 and T_2 are alternating chains,

$$x^0(D) = x^0_{v_{2l}} + k - 1 = x^0_{v_{2l}} + \tfrac{1}{2}(|D| - 1).$$

As $x^0_{v_{2l}} > 0$ by Lemma 3, $x^0(D) > \frac{1}{2}(|D| - 1)$ which is impossible (Lemma 4).

The cases $l = 1$ and $l = k$ are similar.

Remark. Vertices labelled by $P(b)$ or $P(c)$ induce the same labelling on corresponding vertices of $\bar{G}$ (b and c induce label $+1$ on a). Note that, by Lemma 6, this label is well defined.

Lemma 7. *Let C be an odd hole of $\bar{G}$ such that $x^0(C) = \frac{1}{2}(|C| - 1)$. The number of vertices of C labelled with $+1$ is equal to the number of vertices of C labelled with -1.*

Proof. Let C be an odd hole of $\bar{G}$ such that $x^0(C) = \frac{1}{2}(|C| - 1)$ and let $T = (b, \dots, c)$ be its associated chain in G. Again let us set

$$x^0_b = x^0_a \quad \text{and} \quad x^0_c = x^0_a.$$

Describing T from b to c, let v be the last vertex of T labelled by $P(b)$. Note that $v \in V_2$ and has a label -1.

If T_1 is the subchain of T between b and v and T_3 is the subchain of T between v and c,

$$x^0(T) = x^0(T_1) + x^0(T_3) - x^0_v$$

and

$$x^0(C) = x^0(T_1) - x^0_v - x^0_b + x^0(T_3).$$

If T_1 is not an alternating chain between b and v,

$$x^0(T_1) < x^0_b + x^0_v + \tfrac{1}{2}(|T_1| - 2)$$

Therefore,

$$x^0(C) < x^0(T_3) + \tfrac{1}{2}(|T_1| - 2) \tag{6}$$

On the other hand, there exists an alternating chain T_2 from b to v (Lemma 5) (cf. Fig. 4). Concatenating T_2 and T_3 at v, we get a chain T' from b to c, and $|T'| = |T| - |T_1| + |T_2|$.

If D is the associated odd pseudo-cycle to T', $|D| = |C| - |T_1| + |T_2|$. $x^0(T_2) =$

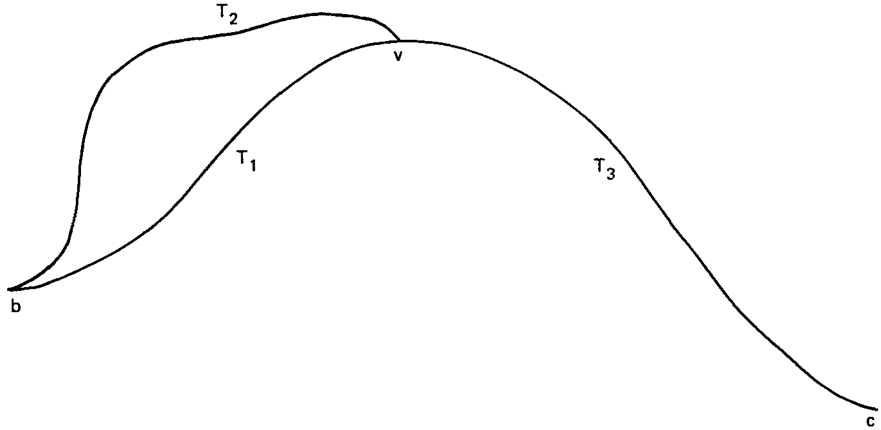


Fig. 4.

$x_b^0 + x_v^0 + \frac{1}{2}(|T_2| - 2)$ since T_2 is an alternating chain. Therefore $x^0(D) = x^0(T_2) - x_b^0 - x_v^0 + x^0(T_3)$.

$$x^0(D) = \tfrac{1}{2}(|T_2| - 2) + x^0(T_3)\,.$$

Using relation (6) we find

$$x^0(D) \geqslant \tfrac{1}{2}(|T_2| - 2) + x^0(C) - \tfrac{1}{2}(|T_1| - 2) = \tfrac{1}{2}(|D| - 1)\,.$$

This is impossible by Lemma 4.

Therefore T_1 is an alternating chain between b and v and the vertices of T_1 are successively labelled with $+1$ and -1 (Lemma 5).

By a similar method we can prove that all the vertices labelled by the Procedure $P(c)$ form a subchain $(w, \ldots, c)$ of T and w has a label $+1$.

All the labelled vertices of C form a chain $(v, \ldots, w)$ such that v has a label -1, w has a label $+1$ and two adjacent labelled vertices have different labels. This implies the result.

3.5. *Proof of Theorem 1*

Let us define the subset of saturated holes $C_1(\bar{G})$ by $C \in C_1(\bar{G})$ if $C \in C(\bar{G})$ and $x^0(C) = \frac{1}{2}(|C| - 1)$.

If x^0 is an extreme point of $P_{\bar{G}}$, x^0 is the unique solution of the linear system

$$(\mathrm{L}) \quad \begin{cases} x_v^0 + x_w^0 = 1 & \forall\ (v, w) \in E_1\,, \\ x^0(C) = \frac{1}{2}(|C| - 1) & \forall\ C \in C_1(\bar{G})\,. \end{cases}$$

Let us define now another solution of system (L): $x^1 = (x_v^1, v \in \bar{V})$ by
- $x_v^1 = x_v^0$ if v is not labelled,
- $x_v^1 = x_v^0 + 1$ if v has a label $+1$,
- $x_v^1 = x_v^0 - 1$ if v has a label -1.

The different steps of Procedure $P(u)$ show that

$$x_v^1 + x_w^1 = 1 \quad \forall\, (v, w) \in E_1 .$$

Moreover, $x^1(C) = \frac{1}{2}(|C| - 1)\ \forall\ C \in C_1(\bar{G})$ by Lemma 7. This proves that x^1 is a solution of linear system (L) distinct from x^0, a contradiction.

3.6. A related result

Let $G = (V_1, V_2, E)$ and $G' = (V_1', V_2', E')$ be two bipartite graphs. Let $\bar{G}$ be the graph obtained by identification of $a \in V_1$ and $a' \in V_1'$ and identification of $b \in V_2$ and $b' \in V_1'$.

Corollary 1. *$\bar{G}$ is a h-perfect graph.*

$\bar{G}$ can be constructed by identifying first a and a'. The graph obtained G' is bipartite. Then we identify b and b' and we apply Theorem 1.

4. Some operations preserving perfectness of graphs

4.1. Theorem 1 cannot be extended to perfect graphs.

Here is a counterexample. Consider graph G, as depicted in Fig. 5. Let us identify $\bar{v}_5$ and $\bar{\bar{v}}_5$ and call v_5 the new vertex in the graph $\bar{G}$. $\bar{G}$ is a well-known graph (the 'Five rays wheel') and it is known that $P(\bar{G})$ is defined by nonnegativity constraints, clique constraints, odd hole constraints and also the constraint

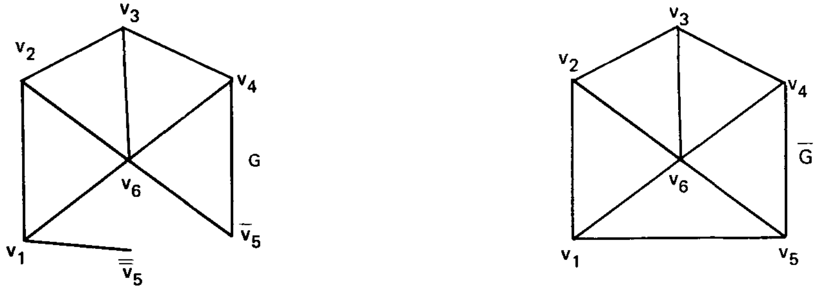


Fig. 5.

$$2x_{v_6}+\sum_{i=1}^{5} x_{v_i} \leq 2 .$$

However, we have the following result.

4.2.

Theorem 2. *Let $G=(V,E)$ be a perfect graph with the following property: there exist two distinct vertices b and c such that all the minimal chains between b and c have odd length. Then, the graph $\bar{G}=(\bar{V},\bar{E})$ obtained by identification of b and c is a perfect graph.*

Proof. (1) Note that b and c are nonadjacent vertices of G. Let us call a the vertex obtained by identification of b and c.

(2) Let us prove that $\Omega \in \Omega(\bar{G})$ if and only if
- $a \notin \Omega$ and $\Omega \in \Omega(G)$,
- or $a \in \Omega$ and $\Omega-\{a\}+\{b\}$ or $\Omega-\{a\}+\{c\}$ belongs to $\Omega(G)$.

It is obvious that a subset of vertices of G which satisfy one of these conditions is a clique of $\bar{G}$.

On the other hand if $\Omega \in \Omega(\bar{G})$ and a does not belong to Ω, Ω is a clique of G. If a belongs to Ω and $\Omega-\{a\}+\{b\}$ is not a clique of G, there exists a vertex $v \in \Omega$ such that $(b, v) \notin E$. Therefore $(c, v) \in E$.

If $\Omega-\{a\}+\{c\}$ is not a clique there exists a vertex w distinct from v such that $(c, w) \notin E$ and $(b, w) \in E$. But this is impossible since the chain (b, w, v, c) is a minimal chain of length 4.

(3) This last result proves that the polytope

$$\bar{P}\quad \begin{cases} x_v \geq 0 & \forall\, v \in \bar{V}, \\ \sum_{v\in\Omega} x_v \leq 1 & \forall\, \Omega \in \Omega(\bar{G}), \end{cases}$$

may also be characterized as the intersection of the polytope $P(G)$ defined by

$$P(G)\quad \begin{cases} x_v \geq 0 & \forall\, v \in V, \\ \sum_{v\in\Omega} x_v \leq 1 & \forall\, \Omega \in \Omega(G), \end{cases}$$

and the hyperplane H of $R^{|V|}$ defined by the equation

$$x_b - x_c = 0 .$$

Elementary results on polytopes show that any extreme point x^0 of $P(G) \cap H$ which is not an extreme point of $P(G)$ may be written as

$$x^0 = \alpha x^1 + \beta x^2, \quad \alpha > 0, \beta > 0, \alpha + \beta = 1$$

where $x^1 = (x_v^1, v \in V)$ is an extreme point of $P(G)$ which satisfies $x_b^1 - x_c^1 > 0$, and $x^2 = (x_v^2, v \in V)$ is an extreme point of $P(G)$ which satisfies $x_b^2 - x_c^2 < 0$.

As x^1 and x^2 are integer extreme points,

$$x_b^1 = 1, \qquad x_c^1 = 0$$

and

$$x_b^2 = 0, \qquad x_c^2 = 1.$$

Since $x_b^0 - x_c^0 = 0$, this implies that

$$\alpha = \beta = \tfrac{1}{2}.$$

Let us set

$$\begin{aligned} V^1 &= (v, v \in V \text{ and } x_v^0 = 1), \\ V^2 &= (v, v \in V \text{ and } x_v^0 = 0), \\ V^3 &= (v, v \in V \text{ and } x_v^0 = \tfrac{1}{2}). \end{aligned} \tag{7}$$

We note that $x_b^0 = x_c^0 = \frac{1}{2}$. Therefore b and c belong to V_3.

(4) Let G' be the subgraph of G induced on the subset of vertices V_3. By (7) there is no clique of cardinality 3 or more in G', and G' is a bipartite graph. Therefore there exist two independent sets S_1' and S_2' which partition V_3 and since all the minimal chains between b and c have odd length we can assume without loss of generality that both b and c belong to S_1'.

By (7) $S_1 = V_1 \cup S_1'$ and $S_2 = V_1 \cup S_2'$ are independent sets of G; therefore x^{S_1} and x^{S_2} belongs to $P(G)$.

Moreover b and c belong to S_1 and not to S_2. Hence x^{S_1} and x^{S_2} belong to H.

But $x^0 = \frac{1}{2}(x^{S_1} + x^{S_2})$. x^0 cannot be an extreme point of $P(G) \cap H$. All the extreme points of $\bar{P}$ are integer extreme points and $\bar{P} = P(\bar{G})$.

This proves that $\bar{G}$ is a perfect graph.

Corollary 2. *Let G be a perfect graph with the following property: there exist two nonadjacent vertices a and b of G such that all the minimal chains between a and b have even cardinality. The graph $\bar{G}$ obtained by adding to the edges of G a new edge (a, b) is perfect.*

Proof. Consider the graph G' obtained by adding to the vertices of G a new vertex $\bar{b}$ and to the edges of G a new edge $(a, \bar{b})$. G is a perfect graph and all the minimal chains between b and $\bar{b}$ have odd cardinality.

By Theorem 2, the graph $\bar{G}$ obtained by identification of b and $\bar{b}$ is perfect.

Conclusion. Theorem 2 and Corollary 2 may be generalized to a more general class of graphs. Moreover, they have interesting applications in the study of perfect graphs. A forthcoming paper will study these extensions and applications.

References

[1] C. Berge, Graphes et Hypergraphes (Dunod, Paris).
[2] M. Boulala and J.P. Uhry, Polytope des indépendants dans un graphe série parallèle, Discrete Math. 27 (1979) 225–243.
[3] V. Chvatal, On certain polytopes associated with graphs, J. Combin. Theory 18 (1975) 138–154.
[4] L. Lovasz, Normal hypergraphs and the perfect graph conjecture, Discrete Math. 2 (1972) 243–267.
[5] N. Sbihi and J.P. Uhry, Une classe de graphe t-parfait, Rappt. de Recherche, IRMA, U.S.M.G., B.P. 53 X, 38041 Grenoble Cedex.

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 97–120

AN ALGORITHM FOR SUBMODULAR FUNCTIONS ON GRAPHS

András FRANK*

Research Institute for Telecommunication, Budapest, Hungary 1026

A constructive method is described for proving the Edmonds–Giles theorem which yields a good algorithm provided that a fast subroutine is available for minimizing a submodular set function.

The algorithm can be used for finding a maximum weight common independent set of two matroids, for finding a minimum weight covering of directed cuts of a digraph, and, as a new application, for finding a minimum cost k strongly connected orientation of an undirected graph.

As a theoretical consequence of the algorithm, we prove a combinatorial feasibility theorem for Edmonds–Giles polyhedron and then we derive a discrete separation theorem which says, roughly, an integer valued submodular function B and an integer valued supermodular function R can be separated by an integer valued modular function provided that $R \leq B$.

0. Introduction

In [2] Edmonds and Giles have proved a quite general min–max relation for submodular functions on graphs. This result includes such specializations as Hoffman's circulation theorem, Edmonds' polymatroid intersection theorem [1] and the Lucchesi–Younger theorem [15, 16] on directed cuts. Despite this generality, the proof is not too difficult to understand, but it is far from being constructive. One of the purposes of the present paper is to describe an algorithmic proof of the Edmonds–Giles theorem. This proof yields a polynomial bounded algorithm provided that a fast subroutine is available for minimizing a submodular set function. It should be noted that such subroutines indeed exist for the specializations mentioned above.

Recently, Grötschel, Lovász and Schrijver [11] developed a procedure for minimizing an arbitrary submodular function. Their algorithm, which uses the ideas of the ellipsoid method, is a good one. It also implies a rather surprising result, namely, the number of sets X whose value $b(X)$ is explicitly needed

* This paper was written, partly, while the author was visiting the University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

during the algorithm can be bounded above by a polynomial function of n, the cardinality of the ground set. I think it is a great challenge for combinatorial optimization to find a 'proper' combinatorial algorithm for minimizing a submodular function ('proper' means that the algorithm may use integer arithmetic only and no approximation procedure).

Actually, the method of Grötschel et al. is suitable for algorithmically solving the Edmonds–Giles problem itself. Hence in this sense the present algorithm is not the first one. However our method which operates with such classical combinatorial devices as augmenting path, labelling technique etc., also provides a proof for the Edmonds–Giles theorem while the method of Grötschel et al. does not lend itself to such a proof. In fact, their method essentially makes use of the theorem itself.

Since the Edmonds–Giles theorem implies Edmonds' matroid intersection theorem as well as Lucchesi–Younger theorem on the maximum number of edge-disjoint directed cuts, the specializations of our procedure obviously provide algorithms for these cases. These specializations are rather important for their own sake, so it seems to be worthwhile to work out the details and exploit the special advantages for these cases. See [7, 8]. As a further application of the method we shall show how to find the cheapest k-strongly connected orientation of a $2k$-edge-connected undirected graph if the two possible orientations of any edge may have different costs. (The existence of such an orientation was proved by Nash-Williams [17]. See also [5].)

A theoretical consequence of our algorithm is a combinatorial feasibility theorem from which a discrete separation theorem will be derived. This states, roughly, that the integer valued super- and submodular functions r and b can be separated by an integer valued modular function provided that $r \leq b$. This theorem can be considered as a counterpart of the well-known 'continuous' result that a concave and a convex function on a convex, compact set in R^n can be separated by a linear function if the concave function nowhere exceeds the convex one.

Another corollary gives a common generalization of the augmenting circuit theorem from network flow theory and its counterpart in matroid intersection theory [12, 13].

1. Preliminaries

Throughout the paper we work with a finite ground set V of n elements. If $A \subseteq V$, the complement of A is denoted by $\bar{A}$. Sets $A, B \subseteq V$ are *co-disjoint* if $\bar{A}$ and $\bar{B}$ are disjoint. Sets $A, B \subseteq V$ are *intersecting* if none of $A \cap B$, $A - B$, $B - A$ is empty. If, in addition, $A \cup B \neq V$, then A and B are *crossing*. A family $\mathcal{F}$ of subsets of V is *intersecting* (*crossing*) if $A \cap B$, $A \cup B \in \mathcal{F}$ for all intersecting (crossing) members A, B of $\mathcal{F}$. A set function b is *submodular* on A, B if $b(A)+b(B) \geq b(A \cap B)+b(A \cup B)$. If equality holds the function is *modular on* A, B. A function r is *supermodular* if $-r$ is submodular. A set A is called a *$u\bar{v}$-set* if $u \in A$, $v \notin A$.

Let $G=(V, E)$ be a directed graph with n vertices and m arrows. (We use the term 'arrow' rather than directed edge.) Multiple arrows are allowed but loops not. An arrow uv *enters* (*leaves*) $B \subset V$ if B is a $v\bar{u}$-set ($u\bar{v}$-set). For $H \subseteq E$, $\rho_H(B)$ stands for the number of arrows in H entering B.

Set $\rho(B)=\rho_E(B)$ and define $\delta_H(B)=\rho_H(\bar{B})$. For a single element set we use $\rho(v)$ instead of $\rho(\{v\})$.

Often we shall not distinguish between a subset H of E and its incidence vector x. For example, $\rho_x(B)=\rho_H(B)$.

Let $\mathcal{F}'$ be a crossing family of subsets of V and A' be a $(0, \pm 1)$ matrix the rows of which correspond to the members of $\mathcal{F}'$, the columns correspond to the elements of E and

$$a'_{F,e}=\begin{cases}-1 & \text{if } e \text{ leaves } F, \\ +1 & \text{if } e \text{ enters } F, \\ \;\;0 & \text{otherwise}.\end{cases}$$

Let b' be an integer-valued function on $\mathcal{F}'$ submodular on crossing members of $\mathcal{F}'$. Without loss of generality we can assume that $V, \emptyset \notin \mathcal{F}'$. Let d be a nonnegative vector in R^E, that is, d is a weighting of the arrows. The theorem of Edmonds and Giles can be formulated as follows.

Theorem 1. *The linear programming problem*

$$\begin{aligned}&\max dx \\ &\text{s.t.}\quad 0 \leq x \leq 1,\ A'x \leq b',\end{aligned} \tag{1}$$

has an integral optimal solution provided that it has a feasible solution at all. If, in addition, d is integer-valued there exists an integral optimal solution to the dual linear programming problem.

Remark. Actually, Edmonds and Giles proved their theorem in a more general form. They allowed d to have negative components and the bounds for x were arbitrary, not necessarily 0 and 1. It should be noted however, that the three special cases mentioned earlier (Edmonds' matroid polyhedron theorem, the Lucchesi–Younger theorem and graph orientation) are consequences of this apparently weaker version. Moreover, if $d \ngeq 0$, the algorithm can simply be modified without losing the polynomial bound. If in (1) the more general constraint $f \leqslant x \leqslant g$ is prescribed (f, g integer-valued) the algorithm can also be extended to handle this case. That transformation assures finite termination and so provides a proof for the general form of the theorem. However in this case the polynomial bound may be destroyed as is the case in network flow theory where the out-of-kilter algorithm is not a good algorithm. In order to get a proper good algorithm for this general case some scaling technique seems to be needed [3]. Here we do not go into details in this direction but only mention how the general problem (in principle) can be converted into form (1).

(a) If $d(e) < 0$, replace $e = uv$ by $e' = vu$. Set $f'(e') = -g(e)$ and $g'(e') = -f(e)$ and $d'(e') = -d(e)$. If $d(e) \geqslant 0$, set $f'(e) = f(e)$, $g'(e) = g(e)$, $d'(e) = d(e)$.

(b) Set $f''(e) = 0$ and $c(e) = g'(e) - f'(e)$ for each arrow e and set $b'(X) = b(X) - \Sigma\,(g'(e)\colon e$ enters $X) + \Sigma\,(f'(e)\colon e$ leaves $X)$ for $X \in \mathcal{F}'$.

(c) Replace any arrow e by $c(e)$ parallel arrows. The new problem is now equivalent to the original one and is of form (1).

First we shall prove Theorem 1 and give an algorithm for the even more special case when the family of subsets in question is intersecting and the function on the family is submodular on each intersecting pair. In Section 8 we show how this proof and algorithm extends to general crossing families.

2. Intersecting families

In order to avoid confusion, instead of $\mathcal{F}'$, let $\mathcal{F}$ denote an intersecting family of subsets of V and assume that $\emptyset \notin \mathcal{F}$, $V \in \mathcal{F}$. Let b be an integer-valued function on $\mathcal{F}$, submodular on intersecting sets and $b(V) = 0$. Let A be defined in the same way as matrix A'. Consider the dual pair of linear programs:

$$\begin{aligned}&\max\ dx\\ &\text{s.t.}\quad 0 \leqslant x \leqslant 1,\ Ax \leqslant b,\end{aligned} \tag{2}$$

and

$$\begin{aligned}&\min\ by + 1z\\ &\text{s.t.}\ \ y, z \geqslant 0,\ (y, z)\begin{bmatrix} A \\ I \end{bmatrix} \geqslant d,\end{aligned} \tag{3}$$

where the components of y and z correspond to the elements of $\mathcal{F}$ and E, respectively, and I denotes the identity matrix (of appropriate size).

The complementary slackness conditions are

$$x(e)>0 \to ya_e + z(e) = d(e)\,,$$
$$z(e)>0 \to x(e) = 1\,,$$
$$y(F)>0 \to a_F^* x = b(F)\,,$$

where a_e denotes the column vector of A belonging to e and a_F^* stands for the row vector of A belonging to F. One can see that y determines uniquely the optimal vector z, namely $z(e) = d(e) - ya_e$, if this value is positive and 0 otherwise. Thus henceforth we shall refer to the dual solution only by the vector y.

Since we are interested in 0–1 vectors x the optimality criteria are as follows:

$$\begin{aligned} &x(e) = 1 \to ya_e \leq d(e)\,,\\ &x(e) = 0 \to ya_e \geq d(e)\,,\\ &y(F)>0 \to a_F^* x = b(F)\,. \end{aligned} \tag{4}$$

Starting with any integral feasible solution to (2), the algorithm will construct vectors x and y satisfying (4). The method is based on two ideas. The first one makes it possible for the dual program to be handled with the help of a so-called potential which is a $|V|$-dimensional vector (unlike the usually much higher dimensional vector y). At the end of the process the optimal y will be simply reconstructed from the potential. The second idea is a generalization of the classical augmenting path method. Roughly speaking, we shall use an augmenting system of disjoint paths rather than one augmenting path for performing a single augmentation. This set of paths will be determined by introducing appropriate auxiliary arrows: a path in the extended graph will define disjoint paths in the original one.

Let us fix a feasible 0–1 solution x to (2) and denote $a_F^* x$ by $\sigma_x(F)$ or briefly $\sigma(F)$. This is a function on $\mathscr{F}$ depending on x and a simple counting argument shows that σ is modular, i.e., $\sigma(K) + \sigma(L) = \sigma(K \cap L) + \sigma(K \cup L)$. The feasibility of x means that $\sigma(F) \leq b(F)$ for each $F \in \mathscr{F}$.

A member F of $\mathscr{F}$ is called *b-strict* (with respect to x) or briefly *strict* when it is not ambiguous if $\sigma(F) = b(F)$. For example, V is always strict since $\sigma(V) = b(V) = 0$.

Lemma 2. *If K and L are intersecting strict members of $\mathscr{F}$, then $K \cap L$ and $K \cup L$ are also strict.*

Proof. $b(K) + b(L) = \sigma(K) + \sigma(L) = \sigma(K \cap L) + \sigma(K \cup L) \leq b(K \cap L) + b(K \cup L) \leq b(K) + b(L)$ from which $\sigma(K \cap L) = b(K \cap L)$ and $\sigma(K \cup L) = b(K \cup L)$ follow, as required. □

Denote by $P(v)$ the intersection of all strict sets containing a vertex v of G. ($P(v)$ depends on x.)

Lemma 3. (a) *$P(v)$ is strict.*

(b) *If a family of strict sets forms a connected hypergraph, the union is again strict.*

Proof. Both statements are direct consequences of Lemma 2. □

3. Potentials

Assume, besides x, we have a vector p in R^V called a *potential* such that

$$x(uv)=1 \to \bar{d}(uv) \geq 0\,, \tag{5a}$$

$$x(uv)=0 \to \bar{d}(uv) \leq 0\,, \tag{5b}$$

$$u \in P(v) \to p(u) \geq p(v)\,, \tag{5c}$$

where $\bar{d}(uv)=d(uv)-p(v)+p(u)$.

Since adding a constant to each component of p does not affect (5) it can be assumed that the minimum component of p is 0. Let the different potential values be $0=p_0<p_1<\cdots<p_k$. If $k>0$ let $V_i=\{u\colon p(u)\geq p_i\}$, $i=1,2,\ldots,k$.

Lemma 4. (5c) *is equivalent to the fact that each V_i partitions into strict sets.*

Proof. For $v \in V_i$, (5c) implies $P(v) \subseteq V_i$, thus the components of hypergraph $\{P(v)\colon v \in V_i\}$ partition V_i. Denote by $\mathcal{K}(V_i)$ the set of these components. Lemma 3(b) states that the members of $\mathcal{K}(V_i)$ are strict.

The reverse direction is obvious. □

The notation $\mathcal{K}(V_i)$ introduced in the proof will also be needed later.

For $F \in \mathcal{F}$, define $y(F)=\Sigma\,(p_i-p_{i-1})$, where the summation is taken over those indices for which $F \in \mathcal{K}(V_i)$. (Here the empty sum is defined to be zero.)

Claim. *For any arrow $e=uv$, $p(v)-p(u)=ya_e$.*

Proof. Let t_i (s_i) denote the number of sets in $\mathcal{K}(V_i)$ which are entered (left) by e. Obviously both t_i and s_i are 0 or 1. Now we have

$$
\begin{aligned}
ya_e &= \sum_F (y(F)\colon e \text{ enters } F) - \sum_F (y(F)\colon e \text{ leaves } F)\\
&= \sum_F \Big(\sum_i (p_i - p_{i-1}\colon F \in \mathcal{K}(V_i))\colon e \text{ enters } F\Big)\\
&\quad - \sum_F \Big(\sum_i (p_i - p_{i-1}\colon F \in \mathcal{K}(V_i))\colon e \text{ leaves } F\Big)\\
&= \sum_{i=1}^k t_i(p_i - p_{i-1}) - \sum_{i=1}^k s_i(p_i - p_{i-1})\\
&= \sum_i (p_i - p_{i-1}\colon e \text{ enters } V_i) - \sum_i (p_i - p_{i-1}\colon e \text{ leaves } V_i)\\
&= \begin{cases} p(v) - p(u) & \text{if } p(v) > p(u),\\ 0 & \text{if } p(v) = p(u),\\ -(p(u) - p(v)) & \text{if } p(v) < p(u). \end{cases}
\end{aligned}
$$

Here we made use of the fact that $t_i = s_i$ whenever $u, v \in V_i$ or $u, v \not\in V_i$. Furthermore, if $p(v) > p(u)$ then the second sum is empty, while $p(u) > p(v)$ implies that the first sum is empty. □

By this claim and the definition of y we need a 0–1 feasible vector x and a potential p satisfying (5).

We shall refer to an arrow uv as a 1-*arrow* (with respect to the given vector x) if $x(uv) = 1$ while uv is a 0-*arrow* if $x(uv) = 0$.

The algorithm will maintain (5a) and (5c) and the number of arrows violating (5b) will gradually reduce.

4. Inner algorithm and proof

The core of our procedure is the following.

Inner algorithm

Input. x : 0–1 feasible solution to (2),
p : potential,
$e = ab$: 0-arrow,

so that (5a) and (5c) hold but e violates (5b).

Output. x': 0–1 feasible solution to (2),
p': potential,

so that (5a) and (5c) continue to hold, e does not violate (5b) and any arrow can violate (5b) only if it violates (5b) with respect to x and p.

Assume this algorithm is available. At the beginning let $p \equiv 0$ and x be an arbitrary 0–1 feasible solution to (2). Repeat the Inner Algorithm until there are no more arrows violating (5b). After no more than $|E|$ applications of this algorithm its output will satisfy all the three criteria in (5).

To describe the Inner Algorithm, define an auxiliary digraph $H = (V, A)$ (depending on the current x and p) as follows. Set $A = A_B \cup A_W \cup A_R$ where

$$A_B = \{uv: uv \text{ is a 1-arrow and } \bar{d}(uv) \leqslant 0\},$$
$$A_W = \{vu: uv \text{ is a 0-arrow and } \bar{d}(uv) \geqslant 0\},$$
$$A_R = \{uv: u \in P(v) \text{ and } p(u) = p(v)\}.$$

(Note that A may contain parallel arrows.) Refer to the elements of A_B, A_W and A_R as blue, white and red arrows, respectively.

Try to find a directed path in H from a to b. There may be two cases.

Case 1. $b \not\in T = \{v: v \text{ can be reached from } a \text{ in } H\}$. Obviously,

(P) there is no arrow in H leaving T.

Revise the potential as follows:

$$p'(v) = \begin{cases} p(v) & \text{if } v \in T, \\ p(v) + \delta & \text{if } v \not\in T, \end{cases}$$

where $\delta = \min(\delta_e, \delta_B, \delta_W, \delta_R)$, where

$$\delta_e = \bar{d}(ab),$$
$$\delta_B = \min\{\bar{d}(uv): uv \text{ is a 1-arrow of } G \text{ leaving } T\},$$
$$\delta_W = \min\{-\bar{d}(uv): uv \text{ is a 0-arrow of } G \text{ entering } T\}$$

and

$$\delta_R = \min\{p(u) - p(v): v \not\in T, u \in P(v) \cap T\}.$$

(Here the minimum is defined to be plus infinity when it is taken over the empty set.)

Claim. $\delta > 0$.

Proof. Since e violated (5b), $\delta_e > 0$. If $\bar{d}(uv) \leqslant 0$ for a 1-arrow uv leaving T, then uv would be a (blue) arrow in H leaving T, contradicting (P), therefore $\delta_B > 0$. If $\bar{d}(uv) \geqslant 0$ for a 0-arrow uv entering T, then vu would be a (white) arrow in H leaving T, contradicting (P), so $\delta_W > 0$. Finally, from (5c), $p(u) \geqslant$

$p(v)$ whenever $u \in P(v)$, that is $\delta_R \geq 0$. If there were u and v with $v \notin T$, $u \in P(v) \cap T$ and $p(u) = p(v)$, then uv would be a (red) arrow in H leaving T, contradicting (P). □

The revised function $\bar{d}'(uv)$ is

$$\bar{d}'(uv) = \begin{cases} \bar{d}(uv) - \delta & \text{if } uv \text{ leaves } T, \\ \bar{d}(uv) + \delta & \text{if } uv \text{ enters } T, \\ \bar{d}(uv) & \text{otherwise}. \end{cases} \tag{6}$$

Claim. (5a) *continues to hold.*

Proof. For a 1-arrow uv, $\bar{d}(uv) \geq 0$. If, indirectly $\bar{d}'(uv) < 0$, then uv leaves T by (6). Now $\bar{d}(uv) \geq \delta_B \geq \delta$, that is, $\bar{d}'(uv) \geq 0$, a contradiction. □

Claim. *If* (5b) *was true for a* 0*-arrow* uv, *it continues to hold.*

Proof. Since $\bar{d}(uv) \leq 0$, the indirect assumption $\bar{d}'(uv) > 0$ and (6) would imply that uv enters T. Now $-\bar{d}(uv) \geq \delta_W \geq \delta$, that is, $\bar{d}'(uv) = \bar{d}(uv) + \delta \leq 0$, a contradiction. □

Claim. (5c) *continues to hold.*

Proof. Note that $P(v)$ does not depend on the potential change. Let $u \in P(v)$ and suppose indirectly that $p'(u) < p'(v)$. Then $v \notin T$, $u \in P(v) \cap T$, thus $p'(u) = p(u)$ and $p'(v) = p(v) + \delta$. Hence $p(u) - p(v) < \delta$. On the other hand $p(u) - p(v) \geq \delta_R \geq \delta$, a contradiction. □

If $\delta = \delta_e$, the arrow $e = ab$ satisfies (5b), and thus the solutions $x' = x$ and p' satisfy the requirements of the Inner Algorithm.

If $\delta < \delta_e$ then repeat the Inner Algorithm using, as inputs, the same x, the revised potential $p := p'$ and the same arrow $e = ab$ which still violates (5b). Observe that the arrow set induced by T in the new auxiliary digraph H' is the same as it was in H. Moreover, the definition of δ ensures that H' contains at least one arrow leaving T (which is blue, white or red according as δ is equal to δ_B, δ_W or δ_R). This implies that the set $T' = \{v: v \text{ can be reached from } a \text{ in } H'\}$ properly includes T. Consequently, after at most $|V| - 1$ iterations, either the equality $\delta = \delta_e$ will hold or vertex b will be reached from a. This is Case 2.

Case 2. There is a directed path from a to b in H. Let U be a shortest path. (Actually, we shall use only the fact that there is no red 'shortcut' arrow to U, that is, if the vertices of U in order are $a = v_0, v_1, \ldots, v_k = b$, then $v_i v_{i+j}$ ($j \geq 2$) must not be a red arrow.)

Since ba is a white arrow in H, U and ba form a directed circuit in H. This may include blue, white and red arrows. Let E_1 be the set of arrows in G which correspond to the blue or white arrows of that circuit. Define a new vector x' as follows:

$$x'(e) = \begin{cases} 1 - x(e) & \text{if } e \in E_1, \\ x(e) & \text{otherwise}. \end{cases}$$

(That is, a 1-arrow in E_1 becomes a 0-arrow, while a 0-arrow will be a 1-arrow.) We shall prove that x' and $p' := p$ satisfy the requirements of the Inner Algorithm. For a member F of $\mathscr{F}$ let $\rho_r(F)$ ($\delta_r(F)$) stand for the number of red arrows of U entering (leaving) F.

Lemma 5. *x' is a feasible solution to* (2).

Proof. The proof consists of proving a number of claims.

Claim. $a_F^* x' = a_F^* x + \rho_r(F) - \delta_r(F)$.

Proof. This is quite easy when $\rho_r(F) = \delta_r(F) = 0$ and, in general, follows by a simple induction on $\rho_r(F) + \delta_r(F)$. □

We have to prove that $a_F^* x' \leq b(F)$. By the claim it suffices to prove that $\rho_r(F) \leq \varepsilon(F)$, where $\varepsilon(F) = b(F) - \sigma_x(F)$ (recall that $\sigma_x(F) = a_F^* x$). Now $\varepsilon(F)$ is submodular on intersecting members of $\mathscr{F}$.

Let uv be a red arrow of U entering F such that $p(u)$ ($= p(v)$) is as large as possible, and if there are more such arrows let uv be the last one on the path U (starting from a).

Claim. $\rho_r(F \cup P(v)) = \rho_r(F) - 1$.

Proof. Since no red arrow enters $P(v)$ and uv does not enter $F \cup P(v)$, $\rho_r(F \cup P(v)) \leq \rho_r(F) - 1$. On the other hand if st is another red arrow of U which enters F, then we claim that $s \not\in P(v)$ (that is, st enters $F \cup P(v)$ as well): in the contrary case $p(s) \geq p(v)$ by (5c) thus the maximal choice of $p(v)$ implies $p(s) = p(t) = p(v)$. However, this implies that sv is a red arrow. Because of the choice of uv, st precedes uv on the path U (starting from a) thus sv is a red shortcut arrow to U, a contradiction. □

Claim. $\rho_r(F) \leq \varepsilon(F)$ *for any* $F \in \mathscr{F}$.

Proof. By induction on $\rho_r(F)$. Observe that "$\varepsilon(F) \geq 0$ for each $F \in \mathscr{F}$" is equivalent to "x is a feasible solution to (2)" and $\varepsilon(F)=0$ if and only if F is strict (with respect to x). Let $\rho_r(F)>0$ and let uv be defined in the same way as in the previous claim. Then

$$\begin{aligned}\varepsilon(F) &= \varepsilon(F)+\varepsilon(P(v)) \geqslant \varepsilon(F\cap P(v))+\varepsilon(F\cup P(v)) \geqslant 1+\varepsilon(F\cup P(v))\\ &\geqslant 1+\rho_r(F\cup P(v)) = \rho_r(F)\,.\end{aligned}$$

Here we used the submodularity of ε, the induction hypothesis for $F\cup P(v)$ and the previous claim. □

This completes the proof of the lemma. □

After proving the lemma, let us investigate what happened to the optimality criteria. Since ab has become a 1-arrow it does not violate (5b). If uv is a new 1-arrow, then vu was a white arrow in H so $\bar{d}(uv)\geqslant 0$. If uv is a new 0-arrow then uv was a blue arrow in H thus $\bar{d}(uv)\leqslant 0$. That is, (5a) continues to hold and new 0-arrow violating (5b) has not arisen.

Claim. (5c) *holds with respect to* x' *and* p'.

Proof. From Lemma 4 we know that V_i is the union of disjoint strict sets $X_1, X_2, \ldots, X_r$ where each X_i is strict with respect to x. Since no red arrow leaves any strict set and no red arrow enters V_i (for a red arrow uv we had $p(u)=p(v)$) we have $\rho_r(X_i)=\delta_r(X_i)=0$ whence $a^*_{X_i}x'=a^*_{X_i}x$, that is, each X_i is strict with respect to x'. Apply again Lemma 4. □

The current primal solution x is 0–1 valued throughout the algorithm regardless the integrality of the objective function d. If, in addition, d was integral, then the current potential p is also integral throughout the process, and hence so is the dual solution (y, z). These observations complete the proof of Theorem 1, when the set system $\mathscr{F}$ in question is intersecting. □

5. Steps of the algorithm

Before describing the algorithms in detail some remarks are needed about the steps and the number of steps of the algorithm. In order to apply the algorithm we have to be able to determine the set $P(v)$ for each vertex v, in any intermediate stage. To this end suppose we have an oracle which can

(O) decide, for any primal solution x and vertices u, v whether or not there exists a strict $v\bar{u}$-member of $\mathscr{F}$.

A simple argument shows that $P(v)$ consists of those vertices u for which the answer is no. This means that, in constructing the auxiliary digraph H belonging to a given stage of the algorithm, A_R can be defined as $A_R = \{uv: p(u) = p(v)$ and there is no strict $v\bar{u}$-set$\}$.

If oracle (O) is available and its run needs at most g steps, then $P(v)$ can be determined in at most gn steps for a fixed vertex v. For all v this means n^2g steps.

Another part of the algorithm tries to find a directed path from a to b in the auxiliary digraph H. This can be done with a well-known labelling technique. If this is accomplished by a breadth-first search then a shortcut free path will automatically be produced, if it exists. If no path exists from a to b in H, the set of labelled vertices will just be T. The labelling procedure needs at most n^2 steps. Moreover, if $\delta < \delta_e$ occurs during the algorithm and the Inner Algorithm is started again with the same x and $p := p'$, then the labels determined previously may be used again (recall that $T \subset T'$). In this case the new auxiliary digraph arises simply from the old one in such a way that some arrows from T to $\bar{T}$ are added while some arrows from $\bar{T}$ to T are deleted. Therefore the overall complexity of the Inner Algorithm can be bounded by $O(n^2 + n^2g)$.

The Inner Algorithm will be applied at most $|E|$ times. From the optimal primal solution x and potential p the optimal dual solution can be obtained in at most $O(n^3)$ steps since the components of the hypergraph $\{P(v): v \in V_i\}$ can be obtained in $O(n^2)$ steps and we have at most n different V_i's. Consequently, the optimal primal–dual solutions to linear programms (2) and (3) can be obtained in at most $O(mn^2g + n^3)$ steps provided that a starting 0–1 feasible solution to (2) and oracle (O) is available.

In order for (O) to be available we need a subroutine for minimizing a submodular function, namely minimize $\varepsilon(F)$ $(= b(F) - \sigma_x(F))$ over the $v\bar{u}$-members F of $\mathcal{F}$. If the minimum is negative, the current vector x is not feasible, if the minimum is zero, then there exists a strict $v\bar{u}$-set, if the minimum is positive, then $u \in P(v)$.

Algorithm for intersecting $\mathcal{F}$

Input. G: directed graph,
$\mathcal{F}$: intersecting family, $\mathcal{F} \subset 2^V$,
$b: \mathcal{F} \to Z$ integer-valued function, submodular on intersecting pairs,
$d: E \to R^+$ nonnegative objective function,
x: 0–1 feasible solution to (2).

Output. x: optimal 0–1 solution to (2),
y, z: optimal solution to (3), which is integral if d is.

Step 1.

1.0. Determine $P(v)$, for each $v \in V$.

1.1. If every 0-arrow satisfies (5b), the current x is optimal. Go to Step 4.

1.2. Select an arrow $e = ab$ violating (5b).

1.3. Form the auxiliary digraph $H = (V, A)$ and try to find a directed path from a to b by the labelling technique (making use of labels determined but not deleted previously). If a path U exists go to Step 3.

Step 2 (Change in potential).

2.0. Let T be the set of the labelled vertices. Count δ and set $p(v) := p(v) + \delta$ whenever $v \notin T$.

2.1. If $\delta = \delta_e$ delete all the labels and go to 1.1.

2.2. Go to 1.3.

Step 3 (Change in x). Denoting by E_1 the set of arrows of G corresponding to the blue and white arrows of the circuit $U + ba$, set

$$x(e) := \begin{cases} 1 - x(e) & \text{if } e \in E_1\,, \\ x(e) & \text{otherwise}\,. \end{cases}$$

Go to 1.0.

Step 4 (Forming the optimal solution (y, z) to (3)).

4.0. Let the different values of p be $0 = p_0 < p_1 < \cdots < p_k$. Set $V_i = \{u\colon p(u) \geq p_i\}$ for $i = 1, 2, \ldots, k$.

4.1. For each i, determine the components of the hypergraph $\{P(u)\colon u \in V_i\}$. Denote by $\mathcal{K}(V_i)$ the set of these components.

4.2. For $F \in \mathcal{F}$, set $y(F) = \sum (p_i - p_{i-1})$, where the summation is taken over those indices i for which $F \in \mathcal{K}(V_i)$. (The empty sum is zero.)

4.3. For a 1-arrow e, set $z(e) = d(e) - ya_e$, for a 0-arrow e set $z(e) = 0$. Halt.

6. Starting feasible solution

In this section we investigate the problem of finding a 0–1 feasible solution to (2). It is assumed again that $\mathcal{F}$ is an intersecting family and b is submodular on intersecting members of $\mathcal{F}$.

Feasibility Theorem. *There exists a* 0–1 *feasible solution to* (2) *if and only if*

$$\sum_i b(X_i) \geq -\delta(\textstyle\bigcup X_i)\,, \tag{7}$$

for disjoint members $X_1, X_2, \ldots, X_k$ of $\mathcal{F}$.

Proof. *Necessity.* For a feasible solution x we have $\Sigma\, b(X_i) \geq \Sigma\, \sigma(X_i) = \sigma(\bigcup X_i) \geq -\delta(\bigcup X_i)$.

Sufficiency. A simple trick due to Hoffman [18] will enable us to reduce the problem to that investigated before. Extend the graph $G = (V, E)$ by adding a new vertex r and $|E|$ new arrows as follows. For each vertex $v \in V$ join $\delta(v)$ parallel arrows from r to v. For each $F \in \mathscr{F}$ let $b'(F) = b(F) + \Sigma_{v \in F}\, \delta(v)$. Obviously b' is submodular on intersecting members of $\mathscr{F}$. Furthermore, since $\delta(F) \leq \Sigma_{v \in F}\, \delta(v)$ and $b(F) \geq -\delta(F)$ by (7), it follows that b' is nonnegative.

Let us consider the linear program (2) with respect to the extended graph G' and the new function b' whereas $\mathscr{F}$ remains the same. A simple argument shows that

(A) the original program has a feasible solution if and only if the new program has a feasible solution $x = (x_1, x_2)$ in which $x_2(e) = 1$ for each new arrow e.

Here the components of x_1 and x_2 correspond to the original and new arrows, respectively.

Let the new objective function be $d(e) = 1$ if e is a new arrow and $d(e) = 0$ if e is old. Since $b' \geq 0$ the identically zero vector is an appropriate starting feasible solution. Apply the algorithm with this starting solution. By (A) what we have to prove is that the value of the optimal solution to the new program is just $|E|$. The algorithm provides a primal solution x and a potential p which satisfy (5). Suppose indirectly that $x_2(ru) = 0$ for a new arrow ru. Observe that only the new arrows violated (5b) at the beginning of the algorithm, therefore $p(r) = 0$ throughout the algorithm. Furthermore, $x_2(ru) = 0$ and (5b) imply that $p(u) > 0$. Therefore the set $X = \{v: p(v) > 0\}$ is non-empty. From Lemma 4 X is a disjoint union of some strict sets X_i. That is $X = \bigcup X_i$ and $b'(X_i) = \sigma_x(X_i)$. Moreover, no original 1-arrow enters X and no original 0-arrow leaves X because of (5a) and (5b), respectively. Thus $\Sigma_i\, \sigma_{x_1}(X_i) = -\delta(X)$. Furthermore, since $x_2(ru) = 0$ we have $\sigma_{x_2}(X) < \Sigma\, (\delta(v)\colon v \in X)$. Consequently, $\Sigma_i\, \sigma_x(X_i) = \Sigma_i\, \sigma_{x_1}(X_i) + \Sigma_i\, \sigma_{x_2}(X_i) < -\delta(X) + \Sigma\, (\delta(v)\colon v \in X)$ from which $\Sigma_i\, b(X_i) + \Sigma\, (\delta(v)\colon v \in X) = \Sigma\, b'(X_i) < \Sigma\, (\delta(v)\colon v \in X) - \delta(X)$, that is, $\Sigma_i\, b(X_i) < -\delta(X)$, contradicting the hypothesis of the theorem. $\square$

7. A discrete separation theorem

In this section we shall make use of the simple observation that, for (2) to have a feasible solution, it suffices to require (7) only for those families of disjoint members $X_1, X_2, \ldots, X_k$ of $\mathscr{F}$ where $X_i \cup X_j \in \mathscr{F}$ implies that $b(X_i)+$

$b(X_j) < b(X_i \cup X_j)$ $(i \neq j)$. Indeed, if (7) were not true in general under this weaker restriction, then $\Sigma\, b(X_i) < -\delta(\bigcup X_i)$ for some family $\{X_1, X_2, \ldots, X_k\}$. Let k be as small as possible such that this inequality holds. Now, for some X_i and X_j, say X_1 and X_2, $X_1 \cup X_2 \in \mathcal{F}$ and $b(X_1) + b(X_2) \geq b(X_1 \cup X_2)$, whence $\{X_1 \cup X_2, X_3, \ldots, X_k\}$ would also violate (7). But this family consists of $k-1$ sets only, contradicting the minimality of k.

Let $\mathcal{K}$ be a family of subsets of S closed under union and intersection. Let R and B be two integer-valued functions on $\mathcal{K}$ which are super- and submodular on any two members of $\mathcal{K}$, respectively.

Discrete Separation Theorem. *If $R(X) \leq B(X)$ whenever $X \in \mathcal{K}$, there exists an integer-valued modular function m such that $R(X) \leq m(X) \leq B(X)$ for each $X \in \mathcal{K}$.*

Proof. We can suppose that $\bigcap(X\colon X \in \mathcal{K}) \neq \emptyset$. For otherwise join an extra vertex to each member of $\mathcal{K}$. Let $(S', \mathcal{K}')$ and $(S'', \mathcal{K}'')$ be two copies of $(S, \mathcal{K})$ and join k parallel arrows from any $s' \in S'$ to $s'' \in S''$ and from s'' to s', where k is a big number. Here 'big' means that the outdegree function δ satisfies:

$$-\delta(X') \leq B(X)\,, \tag{8a}$$

$$\delta(X'') \geq R(X)\,, \tag{8b}$$

$$\delta(X' \cup Y'') \geq R(Y) - B(X)\,, \tag{8c}$$

for any $X, Y \in \mathcal{K}$. Obviously, increasing k, $\delta(X'')$ increases since $X \neq \emptyset$ and so does $\delta(X' \cup Y'')$ whenever $X \neq Y$. Thus, for sufficiently large k, (8a), (8b) and (8c) (for $X \neq Y$) will hold. If $X = Y$ then $\delta(X' \cup X'') = 0$ for any k, however $0 \geq R(X) - B(X)$ follows from the hypothesis.

Denoting by E the set of arrows and by $V = S' \cup S''$ we have a directed graph $G = (V, E)$ and a family $\mathcal{F} = \mathcal{K}' \cup \mathcal{K}'' \cup \{V\}$ on its vertices. (Note that the fact $\emptyset \notin \mathcal{F}$ in (2) requires the assumption $\bigcap(X\colon X \in \mathcal{K}) \neq \emptyset$.) Furthermore set $b(X') = B(X)$ for $X' \in \mathcal{K}'$ and $b(X'') = -R(X)$ if $X'' \in \mathcal{K}''$ and $b(V) = 0$. Using the remark done at the beginning of this section, the Feasibility Theorem requires just the truth of (8a), (8b) and (8c). Therefore, by the Feasibility Theorem we have an integer-valued feasible solution x. Let us define $m(K) = \sigma_x(K') = \rho_x(K') - \delta_x(K')$ for $K \in \mathcal{K}$. Then m satisfies the requirements of the theorem. $\square$

Remark. The main content of the Separation Theorem is that the separating modular function is integer-valued. Actually, the existence of a not necessarily integer-valued separating function follows simply from the classical real separation theorem since a submodular (supermodular) function on $\mathscr{K} \subset 2^S$ can be extended to a convex (concave) function on R^S so that the convex function nowhere exceeds the concave one.

In our treatment the Feasibility Theorem—and so the Separation Theorem—was a by-product of a more or less complicated algorithm. Of course there exist simpler proofs of them which do not use arrow-weights. In [6] we proved directly a theorem in terms of orientations of an undirected graph, which is equivalent to the Feasibility Theorem. However in that paper the Separation Theorem was not explicitly mentioned. In [9] we refine the proof of [6] by extending a method of Lawler and Martel [this volume, pp. 189–200] and prove a feasibility theorem for the general case (when $f \leq x \leq g$). Hence we have a good algorithm not depending on f and g and this allows us to obtain the separating modular function in polynomial time.

For an instance of applicability of the Discrete Separation Theorem we show how Edmonds' matroid intersection theorem [1] follows from it. An equivalent version of Edmonds' theorem states that two matroids M_1 and M_2 on S, with the same rank r, have a common base if and only if $b_1(X)+b_2(S-X) \geq r$ for any $X \subset S$, where b_i is the rank function of M_i, $i=1,2$. To see the sufficiency, let $\mathscr{K}$ consist of all subsets of S, set $B(X)=b_1(X)$ and $R(X)=r-b_2(S-X)$. Since $B(X) \geq R(X)$, by the Discrete Separation Theorem, an integer-valued modular function m separates $B(X)$ and $R(X)$. It is an easy exercise to check that m is 0–1 valued on the vertices and the set $X=\{x: m(x)=1\}$ is just a common base.

Another easy consequence of our separation result is a theorem on polymatroids due to Giles [10]. Let b_1 and b_2 be two submodular functions on all subsets of S such that $b_i(\emptyset)=0$ and b_i is monotone increasing, that is $b_i(X) \geq b_i(Y)$ for $X \supseteq Y$, $i=1,2$.

Theorem 7. *If $x \geq 0$ is an integer-valued vector ($x \in Z^V$) such that $x(T) \leq b_1(T)+b_2(T)$ for each $T \subseteq S$, then $x=x_1+x_2$ for some nonnegative integer-valued vectors x_1 and x_2 for which $x_i(T) \leq b_i(T)$ for each $T \subset S$ and $i=1,2$. (Here $x(T)$ stands for Σ $(x(s)$: $s \in T)$.)*

Proof. Apply the Discrete Separation Theorem to the functions $R(T)=x(T)-b_1'(T)$ and $B(T)=b_2(T)$ where $b_1'(T)=\min_{X\subseteq T}(b_1(X)+x(T-X))$. $\square$

8. Crossing families

In this section we prove the Edmonds–Giles theorem for the more general case of crossing families and show how the algorithm of Section 5 can be extended. The idea behind this extension is that, with a crossing family $\mathcal{F}'$ and function b' on $\mathcal{F}'$ submodular on crossing members of $\mathcal{F}'$, one may associate an intersecting family $\mathcal{F}$ and a function b on $\mathcal{F}$ submodular on intersecting members of $\mathcal{F}$ so that the sets of feasible solutions to (1) and (2) coincide. Then we can apply the algorithm developed for solving (2).

We shall need a theorem due to Lovász [14].

Theorem 8. *Let $\mathcal{F}'' \subset 2^V$ be a crossing family ($\emptyset, V \notin \mathcal{F}''$), b'' be a function on $\mathcal{F}''$ submodular on any two crossing members of $\mathcal{F}''$. Define $\mathcal{F} = \{X\colon X = \bigcup X_i \neq V, X_i \in \mathcal{F}'', X_i \cap X_j = \emptyset\}$ and $b(X) = \min(\Sigma\, b''(X_i)\colon X_i \in \mathcal{F}', X = \bigcup X_i, X_i \cap X_j = \emptyset)$. If $X, Y \in \mathcal{F}$ and $X \cup Y \neq V$, then $X \cup Y$, $X \cap Y \in \mathcal{F}$ and $b(X) + b(Y) \geqslant b(X \cup Y) + b(X \cap Y)$. (Note that $\mathcal{F}'' \subseteq \mathcal{F}$ and $b''(X) \geqslant b(X)$ for $X \in \mathcal{F}''$.)*

Proof. Let $X, Y \in \mathcal{F}$ be such that $X \cup Y \neq V$. Then $X = \bigcup X_i$, $b(X) = \Sigma\, b''(X_i)$ for some disjoint members X_i of $\mathcal{F}''$ and $Y = \bigcup Y_j$, $b(Y) = \Sigma\, b''(Y_j)$ for some disjoint members Y_j of $\mathcal{F}''$.

If we have some members A_i of $\mathcal{F}''$ which form a connected hypergraph and their union is not V, then this union is in $\mathcal{F}''$. Therefore the components formed by the hypergraph $\{X_i\} \cup \{Y_j\}$ are disjoint members of $\mathcal{F}''$, thus $X \cup Y$ is in $\mathcal{F}$. Furthermore $X \cap Y = \bigcup(X_i \cap Y_j\colon X_i \cap Y_j \neq \emptyset)$ whence $X \cap Y \in \mathcal{F}$.

We need the following lemma.

Lemma 9. *Suppose that the members $A_1, A_2, \ldots, A_l$ and $B_1, B_2, \ldots, B_k$, $k, l \geqslant 1$, of $\mathcal{F}''$ partition A and B, respectively, $\{A_i\} \cup \{B_j\}$ forms a connected hypergraph and $A \cup B \neq V$. Then $A \cup B \in \mathcal{F}''$ and $\Sigma\, b''(A_i) + \Sigma\, b''(B_j) \geqslant b''(A \cup B) + b(A \cap B)$.*

Proof. The first part of the lemma is simple (for a similar observation see Lemma 3). To see the inequality we proceed by induction on $k + l$. The case $k + l = 2$ is obvious so assume $k + l > 2$. Deleting an appropriate edge of a hypergraph, say A_k, the resulting hypergraph remains connected. Now the induction hypothesis holds for $A' = A - A_k$ and B; thus

$$\sum_{i=1}^{k-1} b''(A_i) + \sum_{j=A}^{l} b''(B_j) \geqslant b''(A' \cup B) + b(A' \cap B).$$

Adding $b''(A_k)$ to both sides we get

$$\sum b''(A_i) + \sum b''(B_j) \geqslant b''(A' \cup B) + b''(A_k) + b(A' \cap B).$$

Since $A' \cup B$ and A_k are crossing members of $\mathcal{F}''$ we have

$$b''(A' \cup B) + b''(A_k) \geqslant b''(A' \cup B \cup A_k) + b''((A' \cup B) \cap A_k)$$
$$= b''(A \cup B) + b''(B \cap A_k).$$

Furthermore, $A' \cap B$ and $A_k \cap B$ are disjoint. Thus

$$b''(B \cap A_k) + b(A' \cap B) \geqslant b(A \cap B).$$

From the last three inequalities the lemma follows. □

Let $C_1, C_k, \ldots, C_m$ be the components of the hypergraph $\{X_i\} \cup \{Y_j\}$. Applying Lemma 9 to $A = X \cap C_h$ and $B = Y \cap C_h$, $h = 1, 2, \ldots, m$, we get

$$\begin{aligned} b(X) + b(Y) &= \sum_{h=1}^{m} \Big(\sum_i b''(X_i \cap C_h) + \sum_j b''(Y_j \cap C_h) \Big) \\ &\geqslant \sum_{h=1}^{m} (b''((X \cup Y) \cap C_h) + b(X \cap Y \cap C_h)) \\ &\geqslant b(X \cup Y) + b(X \cap Y). \end{aligned}$$

This completes the proof of Theorem 8. □

In fact, what we need is the following version of Lovász's theorem.

Lemma 10. *Let $\mathcal{F}' \subset 2^V$ be a crossing family, b' be a function on $\mathcal{F}'$ submodular on any two crossing members of $\mathcal{F}'$. Define*

$$\mathcal{F} = \{X: X \neq \emptyset,\ X = \bigcap X_i,\ X_i \in \mathcal{F}',\ \bar{X}_i \cap \bar{X}_j = \emptyset\} \cup \{V\}.$$

In other words $\mathcal{F} - \{V\}$ consists of non-empty sets arising as the intersection of some pairwise co-disjoint members of $\mathcal{F}'$. Let $b(X) = \min(\Sigma\, b'(X_i)$: $X = \bigcap X_i$, $x_i \in \mathcal{F}'$, $\bar{X}_i \cap \bar{X}_j = 0)$ and $b(V) = 0$.

Then $\mathcal{F}$ is an intersecting family and b is submodular on any two intersecting members of $\mathcal{F}$.

Proof. Apply Lovász's theorem for $\mathcal{F}'' = \{X: \bar{X} \in \mathcal{F}'\}$ and $b''(X) = b'(\bar{X})$. □

Denote by P_1 and P_2 the polyhedra defined by (1) and (2) respectively. Now, for (1), b' is the given function on $\mathcal{F}'$ while b and $\mathcal{F}$ for (2) are defined as in Lemma 10.

Lemma 11. $P_1 = P_2$.

Proof. Since $\mathcal{F}' \subseteq \mathcal{F}$ and $b(F') \leqslant b'(F')$ for $F' \in \mathcal{F}'$ we have $P_1 \supseteq P_2$. On the other hand, for a vector x in P_1 and for $F \in \mathcal{F}$ we have

$$b(F) = \sum b'(X_i) \geqslant \sum b(X_i) \geqslant \sum \sigma_x(X_i) = \sigma_x(F) \tag{9}$$

for some $X_i \in \mathcal{F}'$, where $F = \bigcap X_i$ and $\bar{X}_i \cap \bar{X}_j = \emptyset$. (9) shows that $x \in P_2$. $\square$

Lemma 12. *For $F \in \mathcal{F} - \{V\}$ the following statements are equivalent:*
(a) *F is b-strict;*
(b) *F is the intersection of some b'-strict members of $\mathcal{F}'$;*
(c) *F is the intersection of some pairwise co-disjoint b'-strict members of $\mathcal{F}'$.*

Proof. (a)$\to$(c) simply follows from (9). (c)$\to$(b) is trivial. To see (b)$\to$(a), let $F = \bigcap F_i$ where each F_i, $i = 1, 2, \ldots, t$, is a b'-strict member of $\mathcal{F}'$. If among these sets F_i there are two which cross, then they can be replaced by their intersection which is a b'-strict member of $\mathcal{F}'$. Thus if we assume t to be minimal, then the sets F_i are pairwise non-crossing and since their intersection F is non-empty they are pairwise co-disjoint. Thus $b(F) \leqslant \Sigma\, b'(F_i) = \Sigma\, \sigma_x(F_i) = \sigma_x(F)$. From this and (9), $b(F) = \sigma_x(F)$, as required. Note that this replacement operation yields a polynomial procedure. $\square$

Taking into consideration Lemmas 10 and 11, in order to solve (1), it suffices to solve (2) with respect to b. The only difficulty arising from this approach is that of how one can work with the new function b when originally only b' is specified and from an algorithmic point of view the definition of b is rather complicated. Fortunately, we do not need the explicit values of b at all. We have seen that, in order to apply the algorithm of Section 5, only oracle (O) has to be available. The following lemma shows that this is indeed the case provided that the same oracle is available concerning the given $\mathcal{F}'$ and b'.

Lemma 13. *Given $x \in P_1$ $(= P_2)$ and $u, v \in V$, there exists a b-strict $v\bar{u}$-set in $\mathcal{F}$ if and only if there exists a b'-strict $v\bar{u}$-set in $\mathcal{F}'$.*

Proof. Let $F' \in \mathcal{F}'$ be a b'-strict $v\bar{u}$-set. Then $b(F') \leqslant b'(F') = \sigma_x(F') \leqslant b(F')$, i.e., F' is b-strict. Conversely, let $F \in \mathcal{F}$ be a b-strict $v\bar{u}$-set. By Lemma 12, F is the intersection of some b'-strict members of $\mathcal{F}'$. One of them is a $v\bar{u}$-set. $\square$

Lemma 13 shows that, in order to get a primal solution x and a potential which satisfy all three optimality criteria, the algorithm developed for intersecting families can be applied without any change for a crossing family as well. The only difference occurs in Step 4 when the optimal dual solution is formed.

Performing Step 4 we shall have an optimal dual solution y to (3). From this we have to get an optimal solution to the dual of (1). For any $F \in \mathcal{F}$ with $y(F) > 0$, F is a b-strict member of $\mathcal{F}$. Let $y'(X_i) := y(F)$ for each X_i where the sets X_i are pairwise co-disjoint b'-strict members of $\mathcal{F}'$ whose intersection is F (see Lemma 12(c)).

It can immediately be seen that this vector y' is an optimal solution to the dual of (1). At this point the proof of Theorem 1 has been completed. In order to complete the algorithm we must be able to find algorithmically the sets X_i mentioned above. The remainder of this section is devoted to this purpose.

Lemma 14. *A b-strict set F, for which the hypergraph $\{P(u)\colon u \in F\}$ is connected, can be obtained constructively as the intersection of pairwise co-disjoint b'-strict members of $\mathcal{F}'$.*

Proof. Let u be a vertex of F. By Lemma 13, $P(u) = \{v\colon$ there is no b'-strict $u\bar{v}$-set$\}$. With the help of oracle (O) we can produce $P(u)$ as the intersection of some b'-strict members of $\mathcal{F}'$. By Lemma 12 we can algorithmically obtain $P(u)$ as the intersection of pairwise co-disjoint b'-strict members of $\mathcal{F}$.

Now suppose that X, Y are two crossing b-strict members of $\mathcal{F}$ and we have obtained co-disjoint b'-strict members X_i and Y_j of $\mathcal{F}'$ such that $X = \bigcap X_i$ and $Y = \bigcap Y_j$. Then $X \cup Y = \bigcap (Z\colon Z = X_i \cup Y_j,\ X_i$ and Y_j are crossing). Here any set Z is b'-strict thus Lemma 11 applies again. That is, we can get $X \cup Y$ too as the intersection of pairwise co-disjoint b'-strict members of $\mathcal{F}'$. Now Lemma 14 follows since $\{P(u)\colon u \in F\}$ is connected. $\square$

Together with the potential p provided by the algorithm let V_i be defined as in Section 4. Recall that $\mathcal{K}(V_i)$ was the collection of components of the hypergraph $\{P(u)\colon u \in V_i\}$. If $y(F) > 0$, then $F \in \bigcup \mathcal{K}(V_i)$ and apply Lemma 14.

Having finished the algorithmic proof of Theorem 1, we state the corresponding Feasibility Theorem for crossing families. The proof proceeds along the same line as that of the first Feasibility Theorem, so it is omitted.

Feasibility Theorem B. *There exists a 0–1 solution to* (1) *if and only if* $\Sigma\, b'(X_{ij}) \geq -\delta(\bigcup X_i)$ *for disjoint non-empty sets* $X_1, X_2, \ldots, X_k$ *where each* X_i *is the intersection of pairwise co-disjoint members* X_{ij} *of* $\mathcal{F}'$, $j = 1, 2, \ldots, k_i$.

9. Augmenting circuits

A basic result of network flow theory states that a feasible circulation is of minimum cost if and only if it admits no augmenting circuit with negative weight. In matroid theory a similar theorem, concerning two matroids on a weighted ground set, states that a common independent set of k elements is of maximum weight if and only if there is no augmenting circuit with negative weight in an appropriately defined auxiliary digraph (see [12, 13]). Here we show that these theorems are specializations of our more general result. For another general augmenting circuit result, see [19].

Let x be a feasible 0–1 solution to (1). Form a digraph $D = (V, \bar{A})$ depending on x as follows. Set $\bar{A} = \bar{A}_B \cup \bar{A}_W \cup \bar{A}_R$ where

$$\bar{A}_W = \{vu\colon\ uv \text{ is a 0-arrow}\},$$
$$\bar{A}_W = \{vu\colon\ uv \text{ is a 0-arrow}\},$$
$$\bar{A}_R = \{uv\colon \text{there is no } b'\text{-strict } v\bar{u}\text{-set in } \mathcal{F}'\}.$$

Let

$$d'(e) = \begin{cases} d(e) & \text{if } e \in \bar{A}_B, \\ -d(e) & \text{if } e \in \bar{A}_W, \\ 0 & \text{if } e \in \bar{A}_R. \end{cases}$$

Augmenting Circuit Theorem. *An integer valued* 0–1 *solution to* (1) *is optimal if and only if there is no negative circuit in D with respect to the valuation d'.*

Proof. Let x be optimal. Starting with this x, apply the algorithm. We shall get a potential p such that x and p satisfy the optimality criteria. Let C be any circuit in D with vertices $x_1, x_2, \ldots, x_k$. The length $\lambda(C)$ of C is $\sum_{i=1}^{k} d'(x_ix_{i+1})$ (where $x_{k+1} = x_1$). If $x_ix_{i+1} \in \bar{A}_B$, then x_ix_{i+1} is a 1-arrow, thus $d'(x_ix_{i+1}) = d(x_ix_{i+1}) \geq p(x_{i+1}) - p(x_i)$. If $x_ix_{i+1} \in \bar{A}_W$, then $x_{i+1}x_i$ is a 0-arrow, thus $d(x_{i+1}x_i) \leq p(x_i) - p(x_{i+1})$, that is, $d'(x_ix_{i+1}) \geq p(x_{i+1}) - p(x_i)$. Finally, if $x_ix_{i+1} \in \bar{A}_R$, then $p(x_i) \leq p(x_{i+1})$, that is, $d'(x_ix_{i+1}) \geq p(x_{i+1}) - p(x_i)$. Now we have $\lambda(C) = \sum_{i=1}^{k} d'(x_ix_{i+1}) \geq \sum (p(x_{i+1}) - p(x_i)) = 0$.

Conversely, suppose x is not optimal. Again apply the algorithm starting with this x and the identically zero potential as inputs. Performing the algorithm, since x is not optimal, Case 2 will occur sometimes, say when the Inner Algorithm is applied for x, a potential p and a 0-arrow ab. There is a path from a to b in the auxiliary digraph H with vertices $a = x_1, x_2, \ldots, x_k = b$. If x_ix_{i+1} is a blue arrow in H, then $x_ix_{i+1} \in \bar{A}_B$ and $d'(x_ix_{i+1}) = d(x_ix_{i+1}) \leq p(x_{i+1}) - p(x_i)$. If x_ix_{i+1} is a white arrow in H, then $x_{i+1}x_i$ is a 0-arrow. Thus $x_ix_{i+1} \in \bar{A}_W$ and $d(x_{i+1}x_i) \geq p(x_i) - p(x_{i+1})$ whence $d'(x_ix_{i+1}) \leq p(x_{i+1}) - p(x_i)$. If

$x_ix_{i+1} \in A_R$, then $x_ix_{i+1} \in \bar{A}_R$ and $d'(x_ix_{i+1}) = 0$. Finally $ba \in \bar{A}_W$ and $d'(ba) < p(a) - p(b)$ (since ab violated (5b), that is $d(ab) > p(b) - p(a)$). Hence the length $\lambda(C)$ of the circuit $C = x_1, x_2, \ldots, x_k, x_1$ is $\sum_{i=1}^k d'(x_ix_{i+1}) < \sum_{i=1}^k p(x_{i+1}) - p(x_i) = 0$. □

If we consider the more general form of (1) when d is not restricted to be nonnegative and $f \leq x \leq g$ is required, then the same theorem is true provided that the auxiliary graph $D = (V, \bar{A})$ is defined as follows. $\bar{A} = \bar{A}_B \cup \bar{A}_W \cup \bar{A}_R$ where

$$\bar{A}_B = \{uv\colon x(uv) > f(uv)\},$$
$$\bar{A}_W = \{vu\colon x(uv) < g(uv)\},$$
$$\bar{A}_R = \{uv\colon \text{there is no } b'\text{-strict } v\bar{u}\text{-set } \mathcal{F}'\}$$

and the costs are

$$d'(e) = \begin{cases} d(e) & \text{if } e \in \bar{A}_B, \\ -d(e) & \text{if } e \in \bar{A}_W, \\ 0 & \text{if } e \in \bar{A}_R. \end{cases}$$

10. Orientations

In this last section we present a new application of Edmonds–Giles theorem which, somewhat surprisingly, concerns undirected graphs. Let $H = (V, A)$ be an undirected graph. The following theorem is due to Nash-Williams [17] (see also [5]).

Theorem 15. *H has a k-strongly connected orientation if and only if H is $2k$-edge connected.*

(A directed graph is *k-strongly connected* if $\rho(X) \geq k$ for $0 \subset X \subset V$.)

Suppose that the two possible orientations uv and vu of an edge may have different costs $c(uv)$ and $c(vu)$. We are interested in a minimum cost k-strongly connected orientation of H.

By means of $c(uv)$ define a directed graph $G = (V, E)$. Let E consist of those arrows uv for which $c(uv) > c(vu)$ and if $c(uv) = c(vu)$, then one of uv and vu (it does not matter which one) also belongs to E. Furthermore, let $d(uv) = c(uv) - c(vu)$. Then G is an orientation of H with nonnegative costs on its arrows.

Our purpose is to reverse some arrows of G so that the new digraph will be k-strongly connected and the total weight of reoriented arrows will be maximum. Such a reorientation can be described by means of a 0–1 vector x where $x(e)$ is 1 if e is to be reoriented and 0 otherwise. Set $\mathcal{F}' = \{X: 0 \subset X \subset V\}$ and $b'(X) = \rho(X) - k$ where $\rho(X)$ is the indegree function of G.

Consider the linear program (1) for this G, $\mathcal{F}'$ and b' and observe that a 0–1 vector x is a feasible solution to (1) if and only if it defines a k-strongly connected reorientation of G. Therefore our algorithm can be applied if we show that, in this case, oracle (O) is available. This is indeed the case since the oracle requires a subroutine to decide whether or not there exist $k+1$ arrow-disjoint paths from u to v in a directed graph which is a simple flow problem.

By means of a similar transformation we can algorithmically find a minimum cost k-strongly connected orientation of H which satisfies some additional constraints. For example, it can be required that the indegree of any vertex v should satisfy the inequality $f(v) \leq \rho(v) \leq g(v)$ where f and g are given in advance.

References

[1] J. Edmonds, Submodular functions, matroids and certain polyhedra, in: R. Guy, H. Hanani and J. Schönheim, eds., Combinatorial Structures and their Applications (Gordon and Breach, New York, 1970).

[2] J. Edmonds and R. Giles, A min–max relation for submodular functions on graphs, Annals of Discrete Mathematics 4 (North-Holland, Amsterdam, 1979).

[3] J. Edmonds and R.M. Karp, Theoretical improvements in algorithmic efficiency for network flow problems, J. ACM 19 (1972).

[4] R.L. Ford and D.R. Fulkerson, Flows in Networks (Princeton University Press, Princeton, 1962).

[5] A. Frank, On the orientation of graphs, J. Combin. Theory Ser. B 28(3) (1980) 251–261.

[6] A. Frank, On disjoint trees and arborescences, in: L. Lovász and V.T. Sós, eds., Proc. Algebraic Methods in Graph Theory (North-Holland, Amsterdam, 1981) pp. 159–169.

[7] A. Frank, How to make a directed graph strongly connected? Combinatorica 1(2) (1981).

[8] A. Frank, A weighted matroid intersection algorithm, J. Algorithms 2 (1981) 328–336.

[9] A. Frank, Finding feasible vectors of Edmonds-Giles polyhedra, J. Combin. Theory Ser. B, submitted.

[10] R. Giles, Submodular functions, graphs and integer polyhedra, D. Th., University of Waterloo, Waterloo, Ontario, 1975.

[11] M. Grötschel, L. Lovász and L. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, Combinatorica 1(2) 1981.

[12] S. Krogdahl, A combinatorial proof of Lawler's matroid intersection algorithm, unpublished manuscript, 1975.

[13] E.L. Lawler, Combinatorial Optimization: Networks and Matroids (Holt, Rinehart & Winston, New York, 1976).

[14] L. Lovász, Flats in matroids and geometric graphs, in: P.J. Cameron, ed., Combinatorial Surveys: Proc. 6th British Combinatorial Conference (Academic Press, New York, 1977).

[15] L. Lovász, On two minimax theorems in graph theory, J. Combin. Theory Ser. B 21(2) (1976) 96–103.

[16] C. Lucchesi and D. Younger, A minimax theorem for directed graphs, J. London Math. Soc. 17(2) (1978).

[17] C.St.J.A. Nash-Williams, Well-balanced orientations of finite graphs and unobstrusive odd-vertex-pairings, in: Recent Progress in Combinatorics (Academic Press, New York, 1969).

[18] A. Hoffman, Some recent applications of the theory of linear inequalities to extremal combinatorial analysis, Proc. Sympos. Appl. Math. 10 (1960).

[19] R. Burkard and H. Hamacher, Minimal cost flows in regular matroids, Math. Programming Stud. 14 (1981) 32–47.

[20] E.L. Lawler and C.U. Martel, Flow network formulations of polymatroid optimization problems, Annals of Discrete Mathematics 16 (1982) pp. 189–200.

Annals of Discrete Mathematics 16 (1982) 121–127

OPTIMAL SUBTREES AND EXTENSIONS

H. GRÖFLIN

IFOR, Eidgenössische Technische Hochschule, Zürich, Switzerland

Th.M. LIEBLING and A. PRODON

Departement de Mathématiques, École Polytechnique Federale de Lausanne, CH-1000 Lausanne, Switzerland

We consider the problem of finding an optimal family of nested rooted subtrees of a tree. We give a linear algorithm for the associated l.p. for this problem.

A generalization of this problem is that of finding an optimal 'lower closed set' of nodes in an acyclic graph for which polyhedral and polarity characterizations are given. These problems are useful relaxations when solving more complicated sequencing problems.

1. Introduction

Many combinatorial optimization problems arising in practice are being handled by implicit or partial enumeration schemes, their success relying on the thorough knowledge of the polyhedral structure and efficient algorithms of relaxed versions of those problems. This note was thus motivated, for consider the following sequencing problem which up to relaxation of budget constraints models a real world situation.

A tree shaped pipeline network for the transportation of one utility from one central source to a number of potential users is to be set up sequentially over K periods numbered from 1 to K.

Demand has the following structure: if user i can be supplied with the utility starting at time t no later than a given period t_i^b, then this user has a constant demand of d_i units per period starting at period $\max(t, t_i^a)$, t_i^a given, otherwise his demand is lost; that is he recurs to an alternative. Each unit supplied yields a constant profit. There are costs associated with the construction of each link that result in annuities starting the period the costs are incurred, plus a final instalment at period K.

A plan to set up the network—with some links possibly never being built—is to be determined, so as to maximize the net discounted profit over a given planning horizon. This problem may be cast into the form of Section 2 by a judicious choice of graphs and parameters therein.

The tree T of Section 2 is constructed from the given network, by adding to it a hanging node corresponding to each user at the appropriate places. The weights on the edges are given by the sum of the discounted operating costs and profits for the corresponding period, as well as the set up costs. These edge weights may vary from one period to the other as a consequence of the given demand structure, the presence of a discount factor and the set up costs.

2. Optimal nested family of subtrees (NFS)

Given a tree $T=(V,r,\tilde{E})$ with root r and K weight functions $c^k:E\to\mathbb{R}$ on its edges, find a family $J^1\subseteq J^2\subseteq\cdots\subseteq J^K\subseteq\tilde{E}$ such that each J^k spans an r-rooted subtree of T and the weight sum $\Sigma_{k=1}^K c^k(J^k)$ is maximized. Herein $c^k(J^k)=\Sigma\{c_j^k: j\in J^k\}$.

Example. Fig. 1 depicts a case with $K=3$ and $\tilde{E}=\{1,2,3,4\}$. The edges are denoted by numbers in brackets, the other numbers beneath the edges represent the weights c_j^k. The optimal solution, given with thick lines is visibly $J^1=\{1,3\}=J^2$, $J^3=\{1,3,4\}$ and has value 5.

The problem may be viewed as a relaxation and a generalization of the sequencing problem treated in Horn [3], i.e., we do not, as Horn does, require $|J^1|=1$, $|J^k|=|J^{k-1}|+1$, $k=2,\ldots,K$, but do not require any particular structure of the weight functions, either.

Suppose the $|\tilde{E}|=N$ edges are numbered from 1 to N such that if there is a path towards r from edge k to edge j, then $j<k$. Call $s(j)$ the edge immediately following j on the path joining j to the root. Without real loss of generality assume that r has degree one.

Let $t(j)$ be the tail and $h(j)$ be the head of j if T is oriented as an arborescence with in-root r.

Let $x^k=(x_j^k: j\in\tilde{E})\in\mathbb{R}^{\tilde{E}}$ be interpreted as the incidence vector of set J^k, whenever it has (0–1)-components.

With these conventions, the NFS problem may be given the formulation of

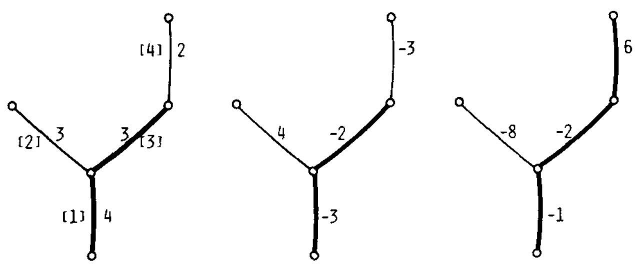


Fig. 1. Example of the NFS problem.

the following linear program:

$$
\begin{aligned}
&\text{maximize } \sum_{j,k} c_j^k x_j^k\,, \\
&\text{s.t. } x_j^k - x_{s(j)}^k \leqslant 0: \quad j = 2, \ldots, N,\ k = 1, \ldots, K\,, \\
&\qquad x_j^k - x_j^{k+1} \leqslant 0: \quad j = 1, \ldots, N,\ k = 1, \ldots, K-1\,, \\
&\qquad x_1^k \leqslant 1: \quad k = 1, \ldots, K\,, \\
&\qquad x_j^k \geqslant 0: \quad j = 1, \ldots, N,\ k = 1, \ldots, K\,.
\end{aligned} \tag{1}
$$

The l.p. (1) has integral extremal solutions due to the fact that it has a totally unimodular constraint matrix, which is immediate from the fact that its dual is the network flow problem (2):

$$
\begin{aligned}
&\text{minimize } \sum \{v_1^k : k = 1, \ldots, K\}\,, \\
&\text{s.t. } -\sum \{v_i^k : s(i) = j\} + v_j^k - h_j^{k-1} + h_j^k \geqslant c_j^k: \quad j = 1, \ldots, N, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 1, \ldots, K\,, \\
&\qquad v_j^k, h_j^k \geqslant 0: \quad j = 1, \ldots, N,\ k = 1, \ldots, K\,, \\
&\qquad (h_j^K \equiv h_j^0 \equiv 0: \quad j = 1, \ldots, N)\,.
\end{aligned} \tag{2}
$$

The network is obtained as follows:

Step 1. Make K copies of the (directed) line graph of arborescence T, where edge j gives rise to node (j, k) of the kth line graph, there is an arc carrying nonnegative flow (v-variables) from node (i, k) to node (j, k), iff $h(i) = t(j)$.

Step 2. For $k = 1, \ldots, K-1$ join (j, k) to $(j, k+1)$ with arcs carrying nonnegative flows (h-variables).

Step 3. Introduce a node 0 and join it to all nodes (j, k) with arcs carrying flow bounded below by the c_j^k.

Step 4. Join the nodes $(1, k)$, $k = 1, \ldots, K$ to node 0 with arcs carrying nonnegative flows (variables v_1^k), whose sum is to be minimized.

The special form of the network makes it possible to formulate an algorithm that solves the dual (2) and the primal (1) in $O(N \cdot K)$ time.

Description of the algorithm

Find the solution to the dual problem by double backward recurrence. First set the flows on the arcs from 0 to (j, k) to their bounds and set the h variables equal to zero. The nodes are then scanned in decreasing order (w.r. to k and w.r. to their distance from the root) as follows: the incoming flows of node (j, k) are known; if their sum is *positive*, set the unique outgoing v-flow equal to this sum and let the incoming h-flow be equal to zero, else set the latter equal to the negative of that sum and the former equal to zero. For the first tree, i.e.,

$k = 1$, use the flows on the arcs between 0 and the nodes $(j, 1)$ instead of the h-variables.

Find the primal solution by double forward recurrence, i.e., a variable x_j^k is set equal to 1 for some k, if its successor $x_{s(j)}^k$ is equal to 1 and the corresponding dual variable v_j^k is positive, in this case, x_j^r will also be set equal to 1 for all $r > k$.

Algorithm

Dual

$z := 0$
for $j := 1, N$ **do**
 for $k := 1, K$ **do**
  begin $h_j^k := 0$; $\tilde{c}_j^k := c_j^k$ **end**
for $k := K$ **downto** 1 **do**
 begin
 for $j := 1, N$ **do begin** $v_j^k := \tilde{c}_j^k$ **end**
 for $j := N$ **downto** 1 **do**
  begin
  if $v_j^k < 0$ **then**
   begin
   if $k > 1$ **then begin** $\tilde{c}_j^{k-1} := \tilde{c}_j^{k-1} + v_j^k$; $\tilde{c}_j^k := \tilde{c}_j^k - v_j^k$; $h_j^{k-1} := -v_j^k$ **end**
   $v_j^k := 0$
   end
  else if $j > 1$ **then** $v_{s(j)}^k := v_{s(j)}^k + v_j^k$
   else $z := z + v_1^k$
  end
 end

Primal

for $j := 1, N$ **do begin** $x_j^0 := 0$ **end**
for $k := 1, K$ **do begin** $x_0^k := 1$ **end**
for $j := 1, N$ **do**
 for $k := 1, K$ **do**
  begin
  if $x_{s(j)}^k = 1$ **and** $(v_j^k > 0$ **or** $x_j^{k-1} = 1)$ **then** $x_j^k := 1$
  else $x_j^k := 0$
  end

Validity of the algorithm

(i) Examination of the dual part of the algorithm shows that it may be viewed as an implementation of a proof by double backward induction (on k and j) of its own validity. In fact it follows immediately that the nonnegative dual variables constructed satisfy the node inequalities of (2) with $=$ for $k>1$ and with $\geqslant$ for $k=1$, hence the solution constructed is dual feasible.

(ii) The primal part obviously produces the incidence vectors of the edge sets of K nested rooted subtrees.

(iii) The value of the dual solution is $z=\sum_{k=1}^{K} v_1^k$ and $v_1^k=\sum_{j=1}^{N} \tilde{c}_j^k x_j^k$, where $\tilde{c}_j^k$ are set at their final values. Thus, to show that the pair of solutions formed by the algorithm is optimal, it suffices to verify that $\sum_{k=1}^{K} \tilde{c}_j^k x_j^k=\sum_{k=1}^{K} c_j^k x_j^k$. But $\tilde{c}_j^k=c_j^k-h_j^k+h_j^{k-1}$ for $k=1,\ldots,K$. For j fixed, let k^* be the smallest index with $x_j^k=1$, if it does not exist, the equation is trivially satisfied, otherwise it is satisfied, if $h_j^{k^*-1}=0$. But $x_j^{k^*}=1$ implies by the definition of k^* and the primal part, that $v_j^{k^*}>0$, hence $h_j^{k^*-1}=0$.

Table 1 gives the optimal solutions of (2) and (1) for the example of Fig. 1.

Table 1

j, k	11	21	31	41	12	22	32	42	13	23	33	43	
c_j^k	4	3	3	2	−3	4	−2	−3	−1	−8	−2	6	
v_j^k	2	0	1	0	0	0	0	0	3	0	4	6	
h_j^k	3	4	2	3	0	8	0	0	0	0	0	0	
x_j^k	1	0	1	0	1	0	1	0	1	0	1	1	$z=5$

3. Optimal lower closed sets of nodes of a graph (LCS)

Let $G=(V,E)$ be a directed, connected acyclic graph and $c: V \to \mathbb{R}$ a weight function on its nodes.

For any $S \subseteq V$, let

$$\delta(S):=\{j \in E:\ t(j) \in S,\ h(j) \notin S\},\quad V_0:=\{v \in V:\ \delta(V-v)=\emptyset\},$$
$$V_T:=\{v \in V:\ \delta(v)=\emptyset\}.$$

We call a subset S of V *lower closed* if, whenever $v \in S$ and $v=t(j)$, then $h(j) \in S$, and denote the family of these sets by $D(G)$, i.e., $D(G):=\{S \subseteq V:\ \delta(S)=\emptyset\}$ ($D(G)$ is the family of node sets corresponding to directed coboundaries, see, for instance, [1]).

The problem of finding $S \in D(G)$ with maximal weight sum can be written as the following l.p.:

$$\text{maximize } c^{\mathrm{T}}x,\ x \in \mathbb{R}^V, \tag{3}$$

$$\text{s.t. } x_v \geqslant 0,\ v \in V_0, \tag{4}$$

$$x_v - x_w \leqslant 0,\ (v,w) \in E, \tag{5}$$

$$x_v \leqslant 1,\ v \in V_T. \tag{6}$$

As in the preceding paragraph, this l.p. can be interpreted as the dual of a network flow problem, and solved as such.

The polyhedron,

$$Q := \{x \in \mathbb{R}^V : x \text{ satisfies (4), (5), (6)}\}, \tag{7}$$

is contained in the unit hypercube, by (4)–(6) and G being acyclic. Its set of extreme points $\text{ext}(Q)$ consists of 0/1 valued vectors and any $x \in \text{ext}(Q)$ is the incidence vector of some set $S \in D(G)$ and vice versa.

The nested subtree problem is an instance of problem (3)–(6), namely for the graph G consisting of K disjoint copies T^k of the directed line graph of $T = (V, \tilde{E})$ and all additional arcs (v^k, v^{k+1}), where v^k and v^{k+1} are copies in T^k and T^{k+1} of $v \in \tilde{E}$, for all $1 \leq k < K$.

Next, we establish a polarity relation between Q and the polyhedron,

$$P = \{x \in \mathbb{R}^V : a^S x \leq 1,\ S \in D(G)\}, \tag{8}$$

where a^S denotes the incidence vector of S. We may write (4), (5) and (6) as

$$\begin{aligned} -e^v x &\leq 0, & v &\in V_0, \\ b^j x &\leq 0, & j &\in E, \\ e^v x &\leq 1, & v &\in V_T, \end{aligned}$$

where $e^v \in \mathbb{R}^v$ denotes the vth unit vector and $b^j \in \mathbb{R}^V$ the vector with components $b^j_{h(j)} = 1$, $b^j_{t(j)} = -1$, $b^j_v = 0$ otherwise, i.e., the jth row vector of the edge-node incidence matrix of G.

Consider

$$P' = \text{conv}\{e^v : v \in V_T\} + \text{cone}\{b^j : j \in E;\ -e^v : v \in V_0\}. \tag{9}$$

We show that $P = P'$.

$P' \subseteq P$ follows immediately from $e^v a^S \leq 1$, $-e^v a^S \leq 0$ and $b^j a^S \leq 0$ for any $v \in V$, $j \in E$, $S \in D(G)$.

Further, using Farkas' lemma, it is easily established, that $P' = \{x \in \mathbb{R}^v : zx \leq 1 \text{ for all } z \in Q\} = \{x \in \mathbb{R}^v : zx \leq 1 \text{ for all } z \in \text{ext}(Q)\}$ and from $\text{ext}(Q) \subseteq \{a^S : S \in D(G)\}$ follows $P \subseteq P'$.

Note that for any directed acyclic graph G there is a unique minimal subgraph G^* in the sense of arc inclusion that preserves the partial order on the nodes induced by G. From this it follows immediately, that the extreme rays of the cone in (9) are defined by the vectors b^j corresponding to the arcs of G^* and $-e^v$, $v \in V^0$.

4. Rooted tree intersections (RTI)

The following is a special case of the LCS problem and a generalization of the NFS problem, as one easily verifies.

Let E be a finite set with weights $c: E \to \mathbb{R}$ associated to its elements. Let further $T_m = (V_m, r_m, E)$, $m = 1, \ldots, M$, be trees with roots r_m, each having E as its set of edges. Without real loss of generality assume r_m is the hanging node of edge $e_0 \in E$ for all $m = 1, \ldots, M$.

Find a subset $J \subseteq E$ that maximizes $c(J) = \Sigma\{c_j : j \in J\}$ and spans a subtree with root r_m in every one of the trees T_m, $1 \leqslant m \leqslant M$.

Let Q_m be the convex hull of the edge incidence vectors of the rooted subtrees of T_m, $1 \leqslant m \leqslant M$. Then we have for the sets of extreme points

$$\operatorname{ext}\left(\bigcap_{m=1}^{M} Q_m\right) = \bigcap_{m=1}^{M} \operatorname{ext}(Q_m).$$

This is in contrast to the case treated in [2], where the rootedness of the subtrees is not required, for consider the following example.

Example. Let $E = \{1, 2, 3, 4\}$ and $M = 2$ with T_1 and T_2 the chains shown in Fig. 2. Let $\mathfrak{A}_m$ be the convex hull of the edge incidence vectors of the subtrees (chains) of T_m. Then, an enumeration of the extreme points of $\mathfrak{A}_1 \cap \mathfrak{A}_2$ shows that this polyhedron has the fractional extreme points

$$(\tfrac{2}{3}, \tfrac{2}{3}, 0, \tfrac{1}{3}), \quad (\tfrac{1}{3}, 0, \tfrac{2}{3}, \tfrac{2}{3}), \quad (0, \tfrac{2}{3}, \tfrac{1}{3}, \tfrac{2}{3}), \quad (\tfrac{2}{3}, \tfrac{1}{3}, \tfrac{2}{3}, 0),$$

aside from the edge incidence vectors of the common subchains of T_1 and T_2, as kindly pointed out to us by W. Altherr.

Fig. 2.

References

[1] J. Edmonds and R. Giles, A min–max relation for submodular functions on graphs, Studies in Integer Programming, Ann. Discrete Math. 1 (1977) 185–204.

[2] H. Gröflin and Th.M. Liebling, Connected and alternating vectors: polyhedra and algorithms, Math. Programming 20 (1981) 233–244.

[3] W.A. Horn, Single-machine job sequencing with tree like precedence ordering and linear delay penalties, SIAM J. Appl. Math. 23(2) (1972) 189–202.

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 129–147

CYCLES THROUGH PRESCRIBED AND FORBIDDEN POINT SETS

D.A. HOLTON

University of Melbourne, Melbourne, Australia

M.D. PLUMMER*

Department of Mathematics, Vanderbilt University, Nashville, TN 37235, USA

A graph G has property $C(m^+, n^-)$ if for any choice of $m + n$ points $u_1, \ldots, u_m, v_1, \ldots, v_n$ in G there is a cycle in G which includes all of $u_1, \ldots, u_m$, but none of $v_1, \ldots, v_n$. We discuss the family of implications '$C(m^+, n^-) \rightarrow C(r^+, s^-)$' for various non-negative integral values of m, n, r and s. The general question as to when such implications hold seems quite difficult. We discuss some reductions of these problems and prove the implication $C(n^+, 1^-) \rightarrow C(n+1^+, 0^-)$ valid for $n = 2$, 3, and 4. The Petersen graph shows that this implication fails for $n = 9$.

1. Introduction

Let m and n be non-negative integers and let G be a graph with at least $m + n$ points. We will say G has property $C(m^+, n^-)$ (or simply, G is $C(m^+, n^-)$) if for all sets S of $m + n$ points in G, where $S = \{u_1, \ldots, u_m, v_1, \ldots, v_n\}$ say, G has a cycle which passes through each of $u_1, \ldots, u_m$, but which misses all of $v_1, \ldots, v_n$.

Special cases of $C(m^+, n^-)$ have been studied by many authors. For example, if $p = |V(G)|$, then $C(p^+, 0^-)$ is equivalent to having a Hamiltonian cycle and the literature here is so vast as to defy inclusion in our bibliography. Two excellent references are surveys by Bermond [1] and Bondy [2].

The largest integer n such that G is $C(n^+, 0^-)$ has been called the *cyclability* of G by Chvátal [4]. Cyclability has also been studied by Halin [7], Holton, McKay, Plummer and Thomassen [10], Mesner and Watkins [11], Plummer and Wilson [13], Sallee [14] and Watkins and Mesner [16]. A close variation is treated in a very recent paper of Bondy and Lovász [3].

In this paper we shall concern ourselves with relationships between

* Supported by the University of Melbourne, Universität Bonn (Sonderforschungsbereich 21 (DFG), Institut für Operations Research), and the University Research Council of Vanderbilt University.

$C(m^+, n^-)$ and $C(r^+, s^-)$, given certain relationships among m, n, r, and s. In particular, when does the implication '$C(m^+, n^-) \rightarrow C(r^+, s^-)$' hold ? This seems to be quite difficult to settle completely. We present partial results and indicate (cf. the proof of Theorem 3.3) the difficulties which—currently, at any rate—seem to prevail.

For any terminology not defined in this paper, the reader is referred to the book by Harary [8] or to a previous paper of McKay, Thomassen, and the present authors [10]. Denote the connectivity of G by $\kappa(G)$. Since we are concerned with the existence of cycles of certain types, throughout this paper, we shall assume our graphs to be 2-connected, unless otherwise specified.

Let u and v be any two points in graph G. Throughout this paper we shall have many occasions to talk of paths with endpoints u and v. We shall use the notation $[u, v]$ to denote such a path which contains both endpoints, (u, v) to denote such a path containing neither endpoint and $[u, v)$ and $(u, v]$ will be the corresponding paths containing only one of the two endpoints. In addition, sometimes we shall want to make it clear to the reader that a path $[u, v]$ also includes among its interior points, say, a, b, c, d and e in which case we shall write $[u, a, b, c, d, e, v]$, with similar notation for 'open' and 'half-open' paths.

Suppose u is a point of graph G and there are three openly disjoint paths joining u to some subgraph H of $G - u$ such that these three paths have only their endpoint $\neq u$ lying in H. We shall have occasion to call the three paths a *tripod* at u.

We shall also wish to consider paths joining two different tripods in a configuration, but having only their endpoints in said configuration. Such a path will be called a *jumper*.

2. Elementary relationships

Consider the lattice of Fig. 1. It is clear that all upward arrows of implication hold; i.e., $C(m^+, n + 1^-) \rightarrow C(m^+, n^-)$ holds for all m and n. For ease of reading we have displayed only those in column two of Fig. 1. It is equally clear that all implications of type $C(m^+, n^-) \rightarrow C(m - 1^+, n^-)$ hold. We display these in Fig. 1.

We proceed to study some of the possible horizontal implications. The result below is due independently to Mesner and Watkins [11] and to Halin [7]. Since a direct proof is short, we include it for completeness.

Lemma 2.1. *Suppose* $2 \leqslant n \leqslant p - 1$. *Then* G *is* $C(2^+, n - 2^-)$ *iff* G *is* n*-connected.*

Proof. ($\leftarrow$) Let a, b, $u_1, \ldots, u_{n-2}$ be any n points in G. By Menger's theorem there are at least n openly disjoint (hereafter o.d.) paths in G joining a and b.

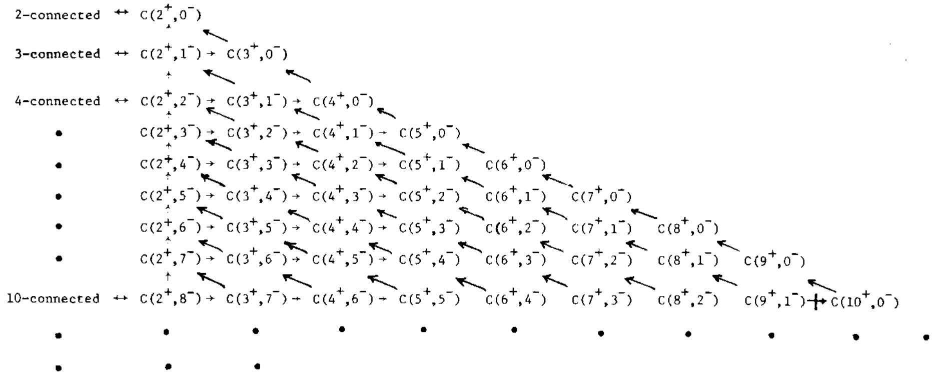


Fig. 1.

But then at least two of these paths—say P_1 and P_2—contain no u_i and the union $P_1 \cup P_2$ is a cycle.

($\rightarrow$) If G is complete, then G is $(p-1)$-connected and hence n-connected. Suppose G is not complete. Then G contains a cutset S of points separating two other points, say a and b. Moreover, we may choose S so that $|S| = \kappa(G) = m$. Now any cycle through a and b must contain at least two points of S and it follows that G is not $C(2^+, m-1^-)$.

Suppose now that G is not n-connected. Thus $\kappa(G) = m \leq n-1$ and hence G is not $C(2^+, n-2^-)$, a contradiction. □

This result establishes the equivalence between the first two columns of Fig. 1. The next lemma due to Halin [7, Satz 5.4] shows that in any horizontal row of Fig. 1, either of the left-hand-most two properties (i.e., k-connected or $C(2^+, k-2^-)$) in fact implies all properties to the right in that particular row. (To avoid clutter we have deleted the corresponding implication arrows in Fig. 1 except those from column 2 toward column 3.)

Lemma 2.2. *Suppose $n \geq 2$. Then G is n-connected iff G is $C(n-k^+, k^-)$ for all k, $0 \leq k \leq n-2$.*

Proof. ($\rightarrow$) Let k be any integer such that $0 \leq k \leq n-2$ and let $u_1, \ldots, u_n$ be any set of n distinct points in G. We must find a cycle through $u_1, \ldots, u_{n-k}$ which misses $u_{n-k+1}, \ldots, u_n$. But $H = G - u_{n-k+1} - \cdots - u_n$ is $(n-k)$-connected, so by a classical theorem of Dirac [6] there is a cycle C in H (and hence in G) through $u_1, \ldots, u_{n-k}$.

($\leftarrow$) Let $k = n-2$ and apply Lemma 2.1. □

Our next lemma shows that if we can verify the top-most left-to-right implication between two contiguous columns of Fig. 1, then in fact all implications directly beneath this one hold.

Lemma 2.3. *Suppose $m \geq 2$ and $n \geq 1$. Then if $C(m^+, n^-) \to C(m+1^+, n-1^-)$ is a valid implication, so is $C(m^+, n+1^-) \to C(m+1^+, n^-)$.*

Proof. Suppose G is $C(m^+, n+1^-)$ and let $U = \{u_1, \ldots, u_{m+1}\}$ and $W = \{w_1, \ldots, w_n\}$ be disjoint subsets of $V(G)$. Now $G - w_n$ is $C(m^+, n^-)$ and hence by hypothesis, $G - w_n$ is $C(m+1^+, n-1^-)$. Hence there is a cycle C in $G - w_n$ through $u_1, \ldots, u_{m+1}$ which avoids $w_1, \ldots, w_{n-1}$. But this same cycle C is a cycle in G passing through all points of U and no points of W and the conclusion follows. □

In particular, by Dirac's theorem if G is 3-connected, then G is $C(3^+, 0^-)$. Combining this with Lemma 3.3 we obtain all left-to-right implications from column 2 to column 3 in Fig. 1.

Our final elementary result relates the size of certain cutsets S in a graph G which is $C(s^+, k^-)$ to the number of components in $G - S$, $c(G - S)$. It will be used many times in proving Theorem 3.3.

Lemma 2.4. *Suppose G is a graph with cutset S, s and k are integers with $s \geq 2$, $k \geq 0$, $|S| = s$ and that G is $C(s^+, k^-)$. Then (1) $s \geq k+2$, and (2) if $k \geq 1$, then $c(G - S) \leq s - k$.*

Proof. (1) Since G is $C(s^+, k^-)$, G is also $C(2^+, k^-)$ and hence $(k+2)$-connected by Lemma 2.1. Thus $s \geq k+2$ as desired.

(2) Suppose $S = \{u_1, \ldots, u_k, u_{k+1}, \ldots, u_s\}$. Also suppose $G - S$ has $s - k + 1$ components, $C_1, \ldots, C_{s-k+1}$ and select a v_i from $V(C_i)$, $1 \leq i \leq s - k + 1$. Since $k \geq 1$, $s - k + 1 \leq s$ and hence G is $C(s - k + 1^+, k^-)$. So let R be a cycle through $v_1, \ldots, v_{s-k+1}$ which misses $u_1, \ldots, u_k$. Then R must use at least $s - k + 1$ points in S whereas, on the other hand, there are only $s - k$ points available in S and we have a contradiction. □

We remark that the demand that $k \geq 1$ is necessary in Lemma 2.4(2), for pick arbitrary integers r and s such that $r \geq s \geq 2$ and consider graph $H = K_s + \bar{K}_r$. (Here the + sign means all possible lines between the complete graph K_s and the independent set $\bar{K}_r$ are present.) Clearly H is $C(s^+, 0^-)$. But, letting $S = V(K_s)$ in H, it is also clear that $G - S$ has r components and, since r may be arbitrarily large, conclusion (2) of Lemma 2.4 fails to hold.

3. Main results

In this section we will show that the implication $C(n^+, 1^-) \to C(n+1^+, 0^-)$ holds for $n = 3$ and 4.

Before proceeding we present a variation on Menger's theorem due to Perfect [12] which will be extremely helpful in proving our two results.

Theorem 3.1. *Let G be a k-connected graph, $a \in V(G)$, $B \subseteq V(G) - \{a\}$ and $|B| \geq k$. Suppose further that there is a set of $k-1$ points $b_1, \ldots, b_{k-1}$ in B and for each $i = 1, \ldots, k-1$ a set of openly disjoint paths P_i joining a to b_i such that $V(P_i) \cap B = \{b_i\}$. Then there exist k openly disjoint paths Q_i joining a to b_i for $i = 1, \ldots, k-1$ and Q_k joins a to a point $b_k \in B - \{b_1, \ldots, b_{k-1}\}$. Moreover, $V(Q_i) \cap B = b_i$ for $i = 1, \ldots, k$.*

It is important to realize at the outset that the Q_i's need not be the same paths as the P_i's.

Theorem 3.2. *If G is $C(3^+, 1^-)$, then G is $C(4^+, 0^-)$.*

Proof. Let a, b, c and d be any four points in G. By hypothesis there is a cycle C through a, b and c, missing d. We seek a cycle through all four of these points. By Lemma 2.2, G is 3-connected, so there are three openly disjoint paths from d to cycle C. Clearly we are done unless the endpoints of these three paths—call them α, β and γ—are pairwise separated on C by points a, b and c (see Fig. 2).

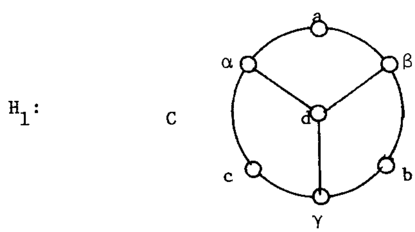


Fig. 2.

Now again by Lemma 2.2 and by Perfect's theorem there are 3 o.d. paths from a to α, β and a point δ where δ lies in the subgraph of H_1 consisting $[\beta, b, \gamma, c, \alpha] \cup [\gamma, d] \cup [\alpha, d] \cup [d, \beta]$. Again, we are finished unless $\delta = \gamma$.

By symmetry there must be a path joining b to α and openly disjoint from $H_1 - (\beta, d] - (\gamma, b]$ and a third from c to β which is openly disjoint from $H_1 - (\alpha, c] - (\gamma, c]$.

Suppose some two of these three paths have a point in common. Without

H_2:

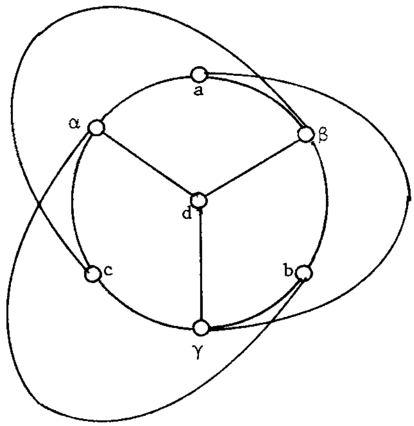


Fig. 3.

loss of generality, suppose $[a, \gamma]$ and $[c, \beta]$ have a point π in common. Then $[a, \pi, c, \gamma, b, \beta, d, \alpha, a]$ is a cycle through a, b, c and d and we are done. So we may suppose all three paths are point-disjoint (cf. Fig. 3).

Thus subgraph H_2 is a homeomorph of the complete bigraph $K_{4,3}$. But since $K_{4,3}$ is not $C(3^+, 1^-)$ there must be a path in $G - H_2$ joining two points of $G - H_2$.

Denote $G - \alpha - \beta - \gamma$ by G_0. By Lemma 2.4, some two of a, b, c and d lie in the same component of G_0. By symmetry we may suppose a and b lie in a common component of G_0. It then follows that there is a path Q in G which joins the tripod at a and that at b, but has only its endpoints in common with $H_2 - \alpha - \beta - \gamma$.

Again by symmetry we may assume the endpoint v_a of Q on the tripod of a lies on $[a, \alpha)$. But then regardless of whether the other endpoint v_b of Q lies on $[b, \alpha)$, $[b, \beta)$ or $[b, \gamma)$, a cycle through a, b, c and d is easily found. □

Theorem 3.3. *If G is $C(4^+, 1^-)$, then G is $C(5^+, 0^-)$.*

Proof. Let a, b, c, d and e be any five points in G. We must prove the existence of a cycle passing through all five. In the many cases to follow we finish by producing such a cycle. For the sake of brevity, we shall refer to such a cycle as a '5-gon'.

By hypothesis, there is a cycle C through b, c, d and e, but which fails to pass through a. Since G is 3-connected by Lemma 2.2, there are three o.d. paths from a to three different points of cycle C, say π_1, π_2 and π_3.

Moreover, we may assume that any two of the π_i's are separated on C by points in $\{b, c, d, e\}$, for otherwise we would be done.

By symmetry, it will suffice to treat the two configurations of Fig. 4.

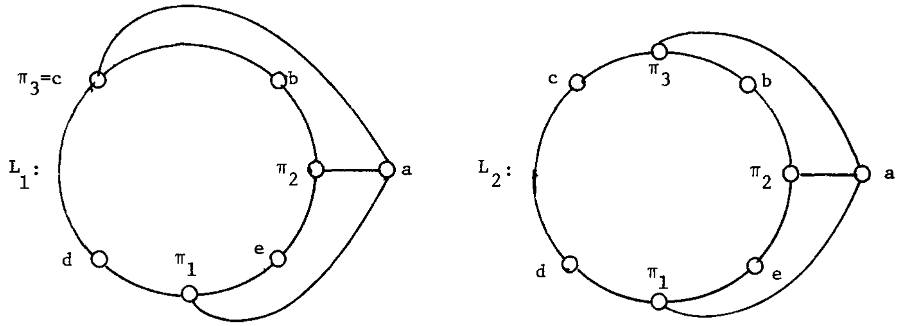


Fig. 4.

First consider the configuration L_1 of Fig. 4. We note that in Theorem 3.2 every cycle through $\{a, b, c, d\}$ also passed through $\{\alpha, \beta, \gamma\}$. Hence the arguments of that theorem show that $\{a, b, c, d, \alpha\}$ all lie on a cycle. Hence the vertices $\{a, b, c, d, e\}$ of the subgraph L_1 also lie on a cycle in G.

Now we assume configuration L_2 of Fig. 4 obtains.

By Perfect's theorem we may assume there are three o.d. paths from b to the rest of the configuration $L_2-(b, \pi_2)-(b, \pi_3)$ and furthermore, that two of these three paths end at points π_2 and π_3. The reader may easily check that there will be a desired 5-gon unless the third path ends at π_1 or at a point π_4 lying in the interior of path $C(c, d)$ (i.e., that part of cycle C between c and d which does not contain b and e) (see Fig. 5).

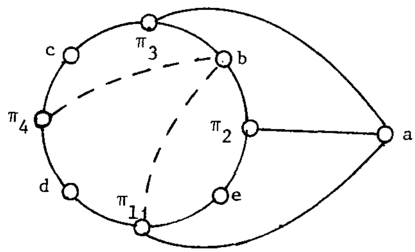


Fig. 5.

Case 1. Suppose the third path joins b and such a π_4. Then by a similar argument from point c, the desired 5-gon is obtained unless the third path from c ends at π_1 or π_2.

Case 1.1. Suppose the third path from c ends at π_1. Then considering three o.d. paths from d, the desired 5-gon is obtained unless the third path from d ends at π_2 or π_3.

Case 1.1.1. Suppose the third path from d ends at π_3. Considering three o.d. paths from e, the third must end in π_3 or π_4 or we are done.

Case 1.1.1.1. Suppose the third path from e ends at π_4. The resulting configuration is thus bipartite (with bipartition $\{\{a, b, c, d, e\}, \{\pi_1, \pi_2, \pi_3, \pi_4\}\}$) (cf. Fig. 6, a drawing reflecting more of the symmetry of the configuration).

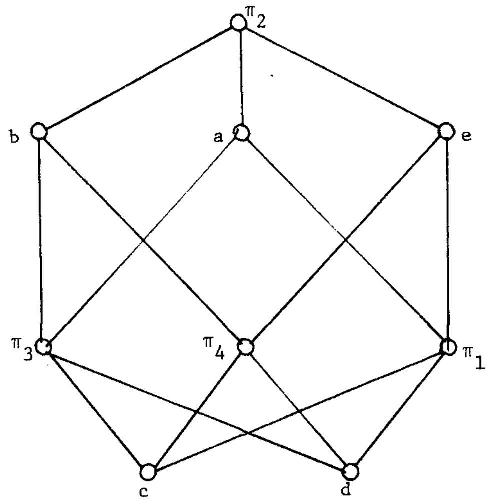


Fig. 6.

One easily sees that if there is a jumper joining a point in the tripod of c to a point in the tripod of d, a desired 5-gon results. Similarly, no jumper from the tripod of c or d to a tripod of a, b or e can exist, except possibly one ending in π_2 (cf. Fig. 7). But $G - \pi_1 - \pi_3 - \pi_4$ has at most two components by Lemma 2.4 and hence a jumper from π_2 to the tripod of c or d must exist. Without loss of generality, assume it joins π_2 to the tripod of c as shown in Fig. 7.

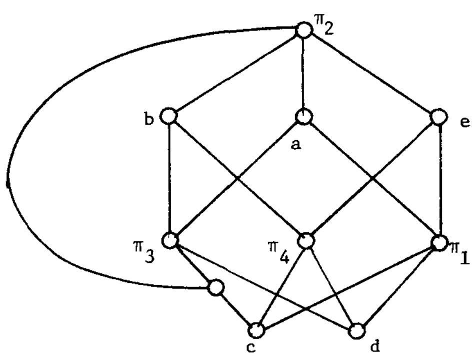


Fig. 7.

Now, by Lemma 2.4, $G-\pi_1-\pi_2-\pi_3-\pi_4$ has at most three components. But c, d and e all lie in different components of $G-\pi_1-\pi_2-\pi_3-\pi_4$ so it follows that $G-\pi_1-\pi_2-\pi_3-\pi_4$ has exactly three components and since a and b cannot share a component with c or d, it follows that a, b and e share the same component of $G-\pi_1-\pi_2-\pi_3-\pi_4$. It follows then that there is a jumper joining some two of the three tripods at a, b and e. By the symmetry, we need treat only one of the three possible cases, say a jumper between the tripods of a and b. But then in G a desired 5-gon results and Case 1.1.1.1 is complete.

Case 1.1.1.2. Suppose the third path from e ends at π_3 (cf. Fig. 8(a)).

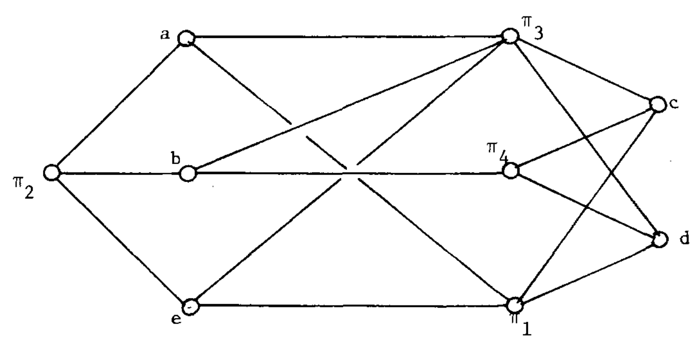


Fig. 8(a).

Again since $G-\pi_1-\pi_2-\pi_3-\pi_4$ can have at most three components, there must be jumpers between some of the tripods at a, b, c, d and e.

It is easy to see that a 5-gon exists for any jumper other than those labelled A, B, C, D or E in Fig. 8(b).

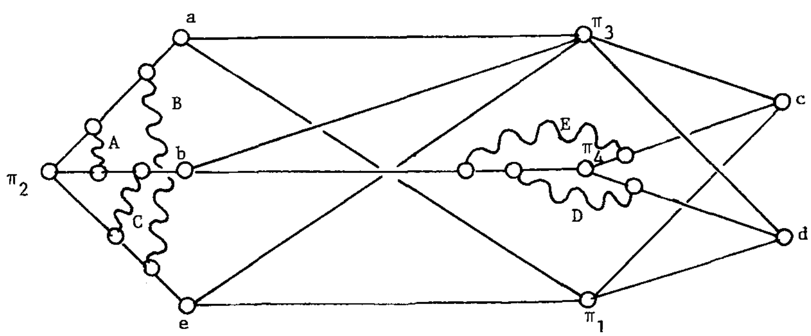


Fig. 8(b).

Claim 1. *There is a jumper of type* D *or* E.

Proof of Claim 1. Suppose not. Now $H = G - \pi_1 - \pi_3 - \pi_4$ has at most two components by Lemma 2.4. First let us suppose that π_2 and c are in the same component of H. Then by our previous work there must be a jumper J from π_2 to the tripod of c. Suppose J ends at a point of $(\pi_3, c]$. Then there are no jumpers of type A, B or C or else we obtain a 5-gon. Thus $G - \pi_1 - \pi_2 - \pi_3 - \pi_4$ has at least four components, contradicting Lemma 2.4. A similar contradiction is reached if J ends on $(\pi_4, c]$ or $(\pi_1, c]$.

Thus π_2 and c are not in the same component of H and similarly neither are π_2 and d. So $\{\pi_1, \pi_3, \pi_4\}$ separates $\{a, b, e\}$ from $\{c, d\}$ in G.

Now there is, by the hypothesis of this theorem, a cycle C in G passing through a, b, c and d, but missing π_3. But then C must pass through π_1 and π_4 and, in fact, $\{\pi_1, \pi_4\}$ separates points c and d from points a and b on cycle C. Let P denote the segment of C joining π_1 and π_4 and containing c and d. Then $P \cup [\pi_4, b] \cup [b, \pi_3] \cup [\pi_3, a] \cup [a, e] \cup [e, \pi_1]$ is a 5-gon and we are finished. This completes the proof of Claim 1 and hence there must be jumpers of type D or E (or both).

Among all jumpers of type D or E, let α_0 be that endpoint of such a jumper which is closest to b on $[b, \pi_4)$. By symmetry we may assume a jumper of type D exists. Let one such be J_D. Note that $\alpha_0 \neq b$ or we are done.

Now there may exist a jumper J joining some point γ of $(\alpha_0, \pi_4]$ to a point β on (b, α_0). Among all such β's choose one—β_0—closest to b on (b, α_0). Again note $\beta_0 \neq b$, or a desired 5-gon exists.

First, let us assume such a J exists (cf. Fig. 9).

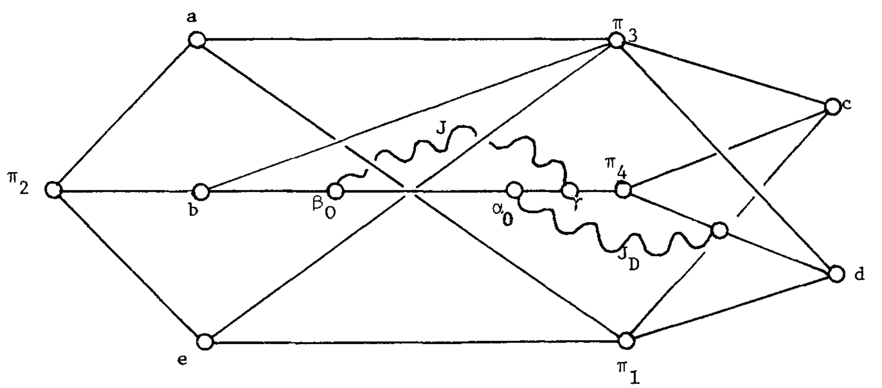


Fig. 9.

Claim 2. *$\{\pi_1, \pi_3, \beta_0\}$ separate $\{a, b, e\}$ from $\{c, d\}$ or the desired 5-gon exists.*

Proof of Claim 2. Suppose $\{\pi_1, \pi_3, \beta_0\}$ does *not* separate $\{a, b, e\}$ from $\{c, d\}$.

Then there must exist a jumper from one of the tripods at a, b or e, or from π_2, to a tripod at c or d or perhaps to the path $(\beta_0, \pi_4] \cup (J - \beta_0) \cup J_D$. Let $\hat{J}$ denote such a jumper. From previous work we know a 5-gon will result unless $\hat{J}$ has one endpoint δ_1 on an a, b or e tripod and the other endpoint δ_2 on $(J - \beta_0) \cup (\beta_0, \pi_4]$.

If δ_2 lies on $(J - \beta_0) \cup (\beta_0, \pi_4)$, a 5-gon is readily found.

So if $\{\pi_1, \pi_3, \beta_0\}$ do not separate $\{a, b, e\}$ from $\{c, d\}$, the only jumpers preventing it either have one endpoint at π_2 or one at π_4 (or both). But then $\hat{H} = G - \pi_1 - \pi_2 - \pi_3 - \pi_4$ certainly has $\{a, b, e\}$ in different component(s) from that (those) of $\{c, d\}$. Moreover, since $\hat{H}$ can have at most three components, a jumper of type A, B or C must exist.

First suppose a type A jumper exists. Then no jumper from π_2 exists or a 5-gon is readily found. So suppose there are jumpers from π_4 (but none from π_2). Then in each case a 5-gon exists and moreover, the said 5-gon does not use the A-type jumper.

Similarly, we are done if B or C-type jumpers exist. This completes the proof of Claim 2.

But then, as before, if $\{\pi_1, \pi_3, \beta_0\}$ separates $\{a, b, e\}$ from $\{c, d\}$ in G, let R denote a cycle containing $\{a, b, c, d\}$ in $G - \pi_3$. R can be modified to produce a 5-gon through $\{a, b, c, d, e\}$.

So now suppose a jumper of type J does not exist.

Claim 3. *$\{\pi_1, \pi_3, \alpha_0\}$ separates $\{a, b, e\}$ from $\{c, d\}$ in G or else a 5-gon exists.*

Proof of Claim 3. Suppose $\{\pi_1, \pi_3, \alpha_0\}$ does not separate $\{a, b, e\}$ from $\{c, d\}$ in G. As before, there must be jumpers from the tripods of a, b or e (or from π_2) to the tripods of c or d or to $(J_D - \alpha_0) \cup (\alpha_0, \pi_4]$ and we are done immediately unless said jumper has one endpoint on $(\alpha_0, \pi_4]$. (N.B. an endpoint on $J_D - \alpha_0$ would be equivalent to a jumper from the tripods of a, b or e to $(\pi_4, d]$ and this has been covered previously.) The proof now proceeds as in Claim 2.

Again, as before, if $\{\pi_1, \pi_3, \alpha_0\}$ separates $\{a, b, e\}$ from $\{c, d\}$ in G, a 5-gon through all of a, b, c, d and e can be found using the hypothesis of the theorem.

This completes Case 1.1.1.2.

Case 1.1.2. Suppose the third path from d ends at π_2. Then a 5-gon is easily obtained unless the third path from e ends at π_3 or π_4.

Case 1.1.2.1. Suppose the third path from e ends at π_3. Then from the drawing of the resulting configuration as rendered in Fig. 10, we see that this case is isomorphic to Case 1.1.1.1.

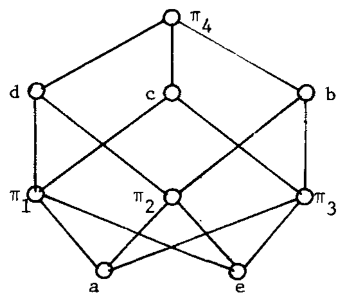


Fig. 10.

Case 1.1.2.2. Suppose then that the third path from e ends at π_4. Then, from the drawing in Fig. 11 of the resulting configuration, once again, we have a case isomorphic to Case 1.1.1.1.

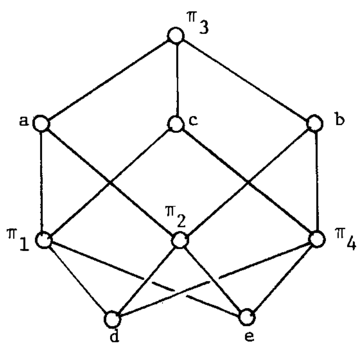


Fig. 11.

Case 1.2. Suppose the third path from c ends at π_2. Applying Perfect's theorem at point d, we obtain a 5-gon unless the third path ends at π_2 or π_3 (see Fig. 12).

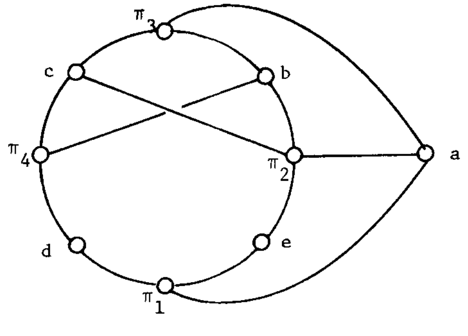


Fig. 12.

Case 1.2.1. If the third path from d ends at π_3, apply Perfect's theorem at e. A 5-gon will result unless the third path ends at π_3 or π_4.

Case 1.2.1.1. Suppose the third path from e ends in π_4. Then the drawing of the resulting configuration in Fig. 13 shows yet again that we have a case isomorphic to Case 1.1.1.1.

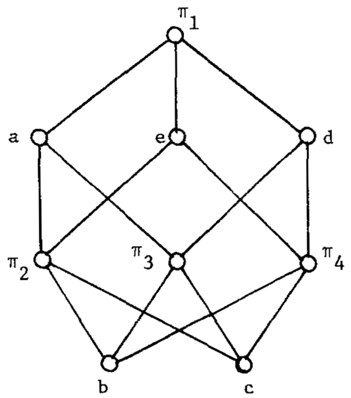


Fig. 13.

Case 1.2.1.2. If the third path from e ends in π_3, then we have the configuration of Fig. 14, which is isomorphic to Case 1.1.1.2.

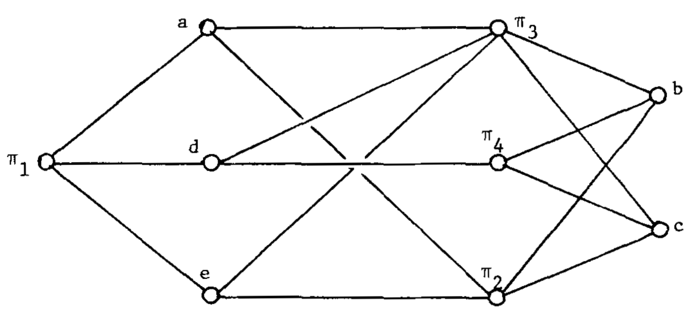


Fig. 14.

Case 1.2.2. Suppose the third path from d ends at π_2. Applying Perfect's theorem at e, a 5-gon results unless the third path ends at π_3 or π_4.

Case 1.2.2.1. Suppose it ends in π_4 (cf. Fig. 15). We see immediately that this case is isomorphic to Case 1.1.1.2.

Case 1.2.2.2. If the third path from e ends at π_3, as before, we see that this case is isomorphic with Case 1.1.1.2.

Case 2. Suppose the third path from b ends in π_1. Applying Perfect's

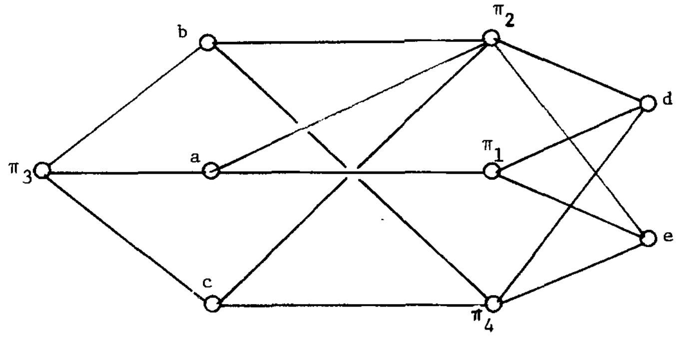


Fig. 15.

theorem at e, we obtain a 5-gon unless the third path ends at π_3, or at a point π_4 in (c, d).

Case 2.1. Suppose it ends in such a π_4. Then redrawing the resulting configuration as in Fig. 16 and comparing with Fig. 12, we see that this case is isomorphic to Case 1.2.

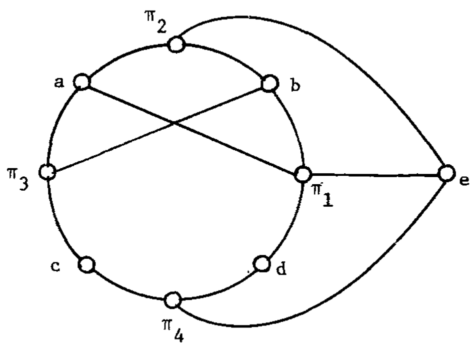


Fig. 16.

Case 2.2. So suppose the third path from e ends at π_3. Since G is 3-connected, we may apply [10] and without loss of generality assume that G contains one of the two configurations in Fig. 17. Furthermore, by this lemma, in Fig. 17(i), $\pi_5 \not\in \{\pi_1, \pi_3\}$ and in Fig. 17(ii), $\pi_7 \not\in \{\pi_1, \pi_3\}$, but $\{x_1, x_2\} \cap \{\pi_1, \pi_3\}$ may be non-empty.

Case 2.2.1. Assume G contains the configuration of Fig. 17(i). Then a 5-gon is present unless $\pi_5 = \pi_2$. In this latter situation, consider the third path from d. We obtain a 5-gon unless it ends at π_2, π_3 or in (c, π_2) (cf. Fig. 18).

Case 2.2.1.1. Suppose the third path from d ends on π_2 (cf. Fig. 19).

By Lemma 2.4, $G - \pi_1 - \pi_2 - \pi_3$ has at most two components. It is easy to see that no jumpers exist between any two of the three tripods at a, b or e, or

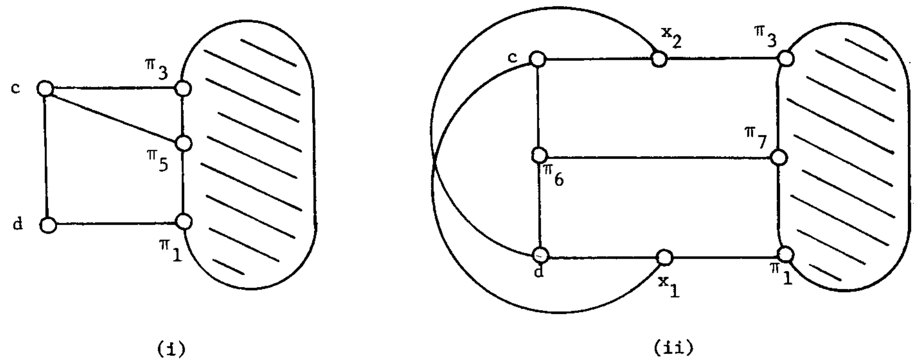


Fig. 17.

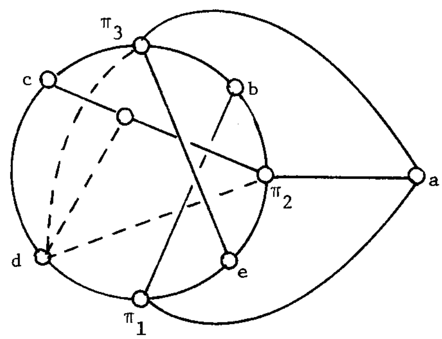


Fig. 18.

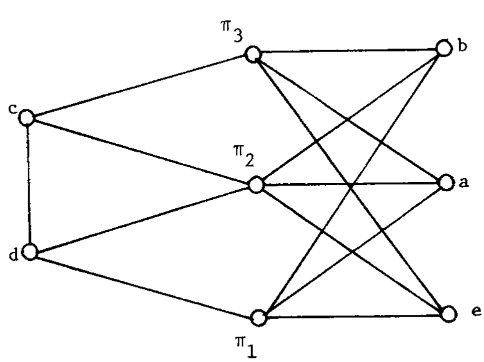


Fig. 19.

else a 5-gon obtains. So by Lemma 2.4 at least two of these three tripods are joined to the $\{c, d\}$-lobe of $G - \pi_1 - \pi_2 - \pi_3$ by jumpers. (Note: by $\{c, d\}$-*lobe* we mean that component of $G - \pi_1 - \pi_2 - \pi_3$ containing c and d.) One easily checks that a 5-gon exists unless there is a jumper J joining a point π_5 in (c, d) to a point on one of the tripods at a, b or e. But if J meets a line of symmetry type $[b, \pi_2)$ we have a configuration isomorphic to Case 1.1.1.1 and if it meets a line of symmetry type $[b, \pi_3)$ the resulting configuration is isomorphic to Case 1.1.1.2.

Case 2.2.1.2. Suppose the third path from d ends on π_3. This case is isomorphic to Case 2.2.1.1.

Case 2.2.1.3. Suppose the third path from d ends at a point π_6 in (c, π_2) (cf. Fig. 20).

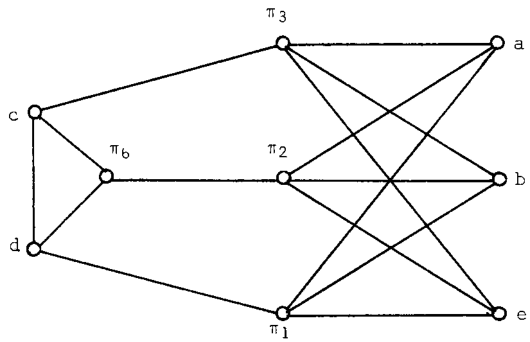


Fig. 20.

Once again, $G - \pi_1 - \pi_2 - \pi_3$ has at most two components, so jumpers exist. If any jumpers join any two of the tripods at a, b and e, we obtain a 5-gon. So by Lemma 2.4 at least two of the tripods at a, b and e are joined to the $\{c, d, \pi_4\}$-lobe of $G - \pi_1 - \pi_2 - \pi_3$. But in all cases here a 5-gon obtains.

Case 2.2.2. Assume G contains the configuration of Fig. 17(ii).

Case 2.2.2.1. Suppose $\{x_1, x_2\} \cap \{\pi_1, \pi_3\} = \emptyset$. By symmetry we need check only six possible locations for point π_7. For example, it suffices to consider $\pi_7 \in (\pi_3, a] \cup (\pi_3, e] \cup (\pi_3, b] \cup [b, \pi_2) \cup (\pi_2, a] \cup \{\pi_2\}$. If π_7 belongs to any of the first five paths, a 5-gon is present. For example, if $\pi_7 \in (\pi_3, a]$, then $[a, \pi_5, \pi_4, c, x_1, d, x_2, \pi_3, e, \pi_1, b, \pi_2, a]$ suffices.

So the sole remaining case to check is $\pi_7 = \pi_2$ (cf. Fig. 21).

Since $G - \pi_1 - \pi_2 - \pi_3$ has at most two components, once again there exist jumpers. But as before any jumper joining any two of the three tripods at a, b and e will help to form a 5-gon. Thus there must exist a jumper from one of these three tripods to the component of $G - \pi_1 - \pi_2 - \pi_3$ containing c and d.

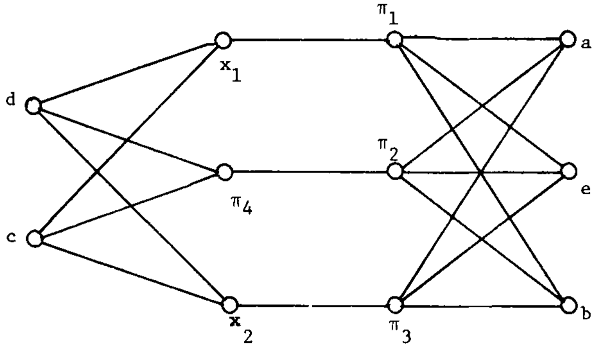


Fig. 21.

But then once again a 5-gon exists in all cases.

Case 2.2.2.2. Suppose $|\{x_1, x_2\} \cap \{\pi_1, \pi_3\}| = 1$. Without loss of generality we may assume that $x_1 = \pi_1$ and $x_2 \neq \pi_3$. Then we have a 5-gon unless $\pi_7 = \pi_2$.

So assume $\pi_7 = \pi_2$.

Since $G - \pi_1 - \pi_2 - \pi_3$ has at most two components, once again there exist jumpers. As before, every jumper joining any two of the three tripods at a, b, and e will help produce a 5-gon. Thus we have a jumper $[\alpha, \beta]$ from one of these three tripods to the $\{c, d\}$-lobe of $G - \pi_1 - \pi_2 - \pi_3$. If $\beta \in [c, d]$, then the 5-gon is formed using the argument above with $\pi_6 = \beta$ and $\pi_7 = \alpha$. If $\beta \in [\pi_6, \pi_2)$, then we replace $[\pi_6, \pi_7] = [\pi_6, \pi_2]$ by $[\pi_6, \beta, \alpha]$ to obtain the 5-gon. If $\beta \in [c, x_2] \cup [d, x_2] \cup [c, \pi_1] \cup [d, \pi_1]$ we are in Case 2.2.1. In the remaining cases, i.e., $\beta \in [x_2, \pi_3]$, a 5-gon is readily found.

Case 2.2.2.3. Suppose $x_1 = \pi_1$ and $x_3 = \pi_3$. Since G is $C(4^+, 1^-)$, there is a cycle through b, a and e missing π_1.

In the configuration M_1 of Fig. 22 this cycle X requires the use of each of the paths $[\pi_3, b]$, $[\pi_3, a]$ and $[\pi_3, e]$. Hence there must exist a jumper from one of these three paths, or from one of the three paths $[\pi_1, b]$, $[\pi_1, a]$ or $[\pi_1, e]$. By

M_1:

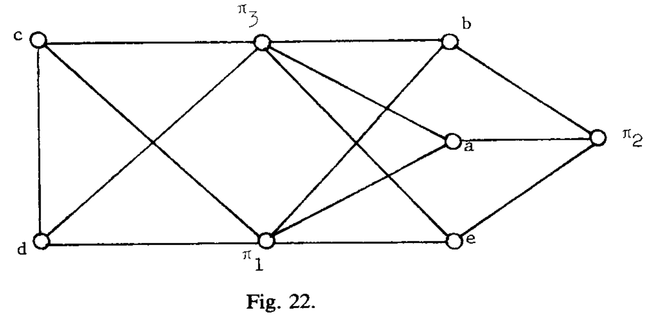


Fig. 22.

symmetry, we may assume a jumper $[\alpha_1, \beta_1]$ from α_1 in (π_3, b). If $\beta_1 \in (\pi_3, a) \cup (\pi_3, e) \cup (\pi_1, a) \cup (\pi_1, e) \cup (\pi_2, a) \cup (\pi_2, e)$, a 5-gon exists.

Suppose $\beta_1 \in (\pi_1, b)$. If $(\alpha_1, \beta_1) \subset X$, then, since $(\pi_1, \beta_1) \not\subset X$, $(\beta_1, b) \subset X$. Similarly, $(\pi_3, \alpha_1) \not\subset X$, so $(\alpha_1, b) \subset X$. So X contains a triangle $[\alpha_1, \beta_1, b, \alpha_1]$, a contradiction.

A similar contradiction results if $\beta_1 \in (\pi_2, b)$.

If $\beta_1 \in [c, \pi_1) \cup [c, \pi_3) \cup [d, \pi_1) \cup [d, \pi_3)$, then the 5-gon is obtained as in Case 2.2.1. Hence $\beta_1 \in (c, d)$ (cf. configuration M_2 of Fig. 23).

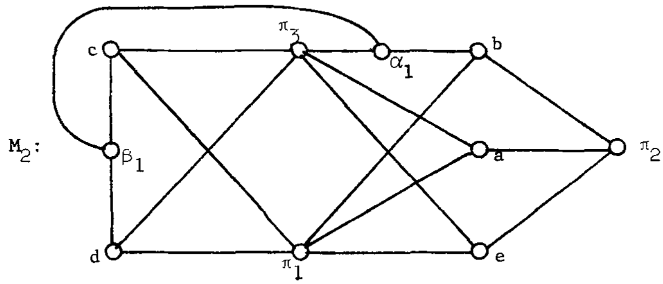


Fig. 23.

But note that if we delete path (π_3, α_1) from M_2 the resulting configuration is precisely that of Fig. 8 and we are done by Case 1. □

We think it worthwhile to point out to the reader that in a 3-connected graph if there is a cycle through some four points which misses a fifth, then it does not necessarily follow that there is a cycle through all five. Just consider the complete bipartite graphs $K(4, n)$ for all $n \geq 5$.

4. Concluding remarks

The implication $C(n^+, 1^-) \rightarrow C(n+1^+, 0^-)$ does not hold for $n = 9$ as is shown by the Petersen graph. In fact, this implication fails for $n = 9$, 12, 14, 15, and for all $n \geq 17$ as is demonstrated by the known existence of hypohamiltonian graphs of order $n + 1$ for each n in this range. For a survey of the state of the art of hypohamiltonian graphs, see [5, 15]. In particular, if there is a counterexample to the above implication for values of $n = 10$, 11 or 13 it cannot be a hypohamiltonian graph, for it has been shown that hypohamiltonian graphs of orders 11, 12 and 14 do not exist. It is unknown at this time whether there is a hypohamiltonian graph on 17 points.

Despite much searching, the only counterexamples to the validity of the implication known to the present authors remain these hypohamiltonian graphs. The Petersen graph is known to be the smallest hypohamiltonian graph. These facts lead us to close by hazarding two conjectures.

Conjecture 1. $C(n^+, 1^-) \to C(n+1^+, 0^-)$ *holds for* $n = 5, 6, 7$ *and* 8.

Conjecture 2. *If* $C(n^+, 1^-) \to C(n+1^+, 0^-)$ *fails, the only exceptions are hypohamiltonian.*

References

[1] J.-C. Bermond, Hamiltonian graphs, in: L.W. Beineke and R.J. Wilson, eds., Selected Topics in Graph Theory (Academic Press, New York, 1979) pp. 127–167.
[2] J.A. Bondy, Hamiltonian cycles in graphs and digraphs, Proc. 9th S.E. Conf. Combinatorics, Graph Theory and Computing (Utilitas Math., Winnipeg, 1979) pp. 3–28.
[3] J.A. Bondy and L. Lovász, Cycles through specified vertices of a graph, Res. Rept. CORR 79-14, Univ. Waterloo (1979).
[4] V. Chvátal, New directions in Hamiltonian graph theory, in: F. Harary, ed., New Directions in the Theory of Graphs (Academic Press, New York, 1973) pp. 65–95.
[5] J.B. Collier and E.F. Schmeichel, Systematic searches for hypohamiltonian graphs, Networks 8 (1978) 193–200.
[6] G.A. Dirac, In abstrakten Graphen vorhandene völlstandige 4-Graphen und ihre Unterteilungen, Math. Nachr. 22 (1960) 61–85.
[7] R. Halin, Zur Theorie der n-fach zusammen-hängenden Graphen, Abh. Math. Sem. Hamburg 33 (1969) 133–164.
[8] F. Harary, Graph Theory (Addison-Wesley, Reading, 1969).
[9] D.A. Holton, B.D. McKay and M.D. Plummer, A corollary to Perfect's theorem, Res. Rept. No. 21, Dept. of Math., Univ. Melbourne, 1981.
[10] D.A. Holton, B.D. McKay, M.D. Plummer and C. Thomassen, Cycles through specified vertices in 3-connected graphs, Combinatorica 1 (1981) 409–418.
[11] D.M. Mesner and M.E. Watkins, Some theorems about n-vertex connected graphs, J. Math. Mech. 16 (1966) 321–326.
[12] H. Perfect, Applications of Menger's graph theorem, J. Math. Anal. Appl. 22 (1968) 96–111.
[13] M.D. Plummer and E.L. Wilson, On cycles and connectivity in planar graphs, Canad. Math. Bull. 16 (1973) 283–288.
[14] G.T. Sallee, Circuits and paths through specified nodes, J. Combin. Theory Ser. B 15 (1973) 32–39.
[15] C. Thomassen, On hypohamiltonian graphs, Discrete Math. 10 (1974) 383–390.
[16] M.E. Watkins and D.M. Mesner, Cycles and connectivity in graphs, Canad. J. Math. 19 (1967) 1319–1328.

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 149–168

AN EFFICIENT PRIMAL SIMPLEX ALGORITHM FOR MAXIMUM WEIGHTED VERTEX PACKING ON BIPARTITE GRAPHS*

Yoshiro IKURA and George L. NEMHAUSER

Department of Operations Research, Upson Hall, Cornell University, Ithaca, NY 14850, USA

We present in graphical terms a primal simplex algorithm for maximum weighted vertex packing on bipartite graphs. For a graph on n vertices, when all of the weights equal one, the number of pivots is bounded by n^2 and the running time of the algorithm is $O(n^4)$. For general integer weights a scaling technique is used and the bounds are increased by a factor equal to the logarithm of the largest weight.

1. Introduction

Let $G=(V,E)$ be a finite, undirected, loopless graph without multiple edges. A *vertex packing, independent set* or *stable set* in G is a subset $P \subseteq V$ such that $v_i, v_j \in P$ implies $(v_i, v_j) \notin E$; a *clique* in G is a subset $C \subseteq V$ such that $v_i, v_j \in C$ implies $(v_i, v_j) \in E$.

Let w_j be a real, positve number assigned to $v_j \in V$; w_j is called the *weight* of v_j. The *maximum weighted vertex packing* problem is to find a vertex packing P such that $w(P) = \Sigma_{j:\, v_j \in P}\, w_j$ is maximum. The case of $w_j = 1$ for all $v_j \in V$ is called the *cardinality* problem.

The weighted vertex packing problem is NP-hard for general graphs. Furthermore, many restricted cases, such as the cardinality problems on triangle-free graphs [14] and cubic planar graphs [7], are also NP-hard. On the other hand, polynomial algorithms are known for claw-free graphs [13, 15], graphs without long odd cycles [9], and perfect graphs [8]. The latter class, which includes bipartite graphs, comparability graphs and chordal graphs, is particularly interesting because of the connection between the combinatorial packing problem and linear programming.

On perfect graphs, the maximum weighted vertex packing problem can be formulated as the linear program [2, 6, 12]:

* This research has been supported by National Science Foundation Grant ECS-8005350 to Cornell University.

$$\max \sum_{j:\, v_j \in V} w_j x_j ,$$

(VLP)

$$\sum_{j:\, v_j \in C} x_j \geqslant 1 \quad \text{for all maximal cliques } C, \tag{1.1}$$

$$x_j \geqslant 0 \quad \text{for all } v_j \in V, \tag{1.2}$$

since there is a one-to-one correspondence between extreme points of the polytope and packings. Although the number of cliques can be exponential in the number of vertices, the ellipsoid method of Khachian [11] has been adapted to run in polynomial time for (VLP) on perfect graphs by Grötschel et al. [8]. Nevertheless, their method is not combinatorial and is unlikely to be practical for computation. The challenge still remains to find efficient combinatorial procedures for the class of perfect graphs. For the subclasses of perfect graphs mentioned above, efficient combinatorial algorithms are known. In particular, the weighted vertex packing problem on bipartite graphs can be reduced to a maximum flow problem.

Here we give an efficient primal 'simplex-like' method in graphical terms for solving (VLP) on bipartite graphs. For the cardinality problem, the number of pivots is bounded by n^2 and the running time of the algorithm is $O(n^4)$, where n is the number of vertices. For general integer weights, the bound on the number of pivots in the basic algorithm involves the sum of the vertex weights; however, using a scaling technique of Edmonds and Karp [5], the bound on the number of pivots is reduced to $n^2(l+1)$, where

$$2^l \leqslant \max_{j:\, v_j \in V} w_j < 2^{l+1} .$$

While no unsolved problems are treated in this paper and the running time bound of our algorithm is not an improvement on existing bounds, we view our contribution as a step in the development of polynomially bounded simplex methods. Furthermore, our approach may be applicable to vertex packing problems on larger classes of perfect graphs.

2. Graphical description of primal and dual solutions

For bipartite graphs, the clique constraints (1.1) simplify to

$$x_i + x_j \leqslant 1 \quad \text{for all } (v_i, v_j) \in E, \tag{2.1}$$

and the dual of (VLP) is the problem of covering weighted vertices by a minimum number of edges:

$$\min \sum_{e\in E} y_e\,,$$

(ELP)
$$\sum_{e:\, v_j\in e} y_e \geqslant w_j \quad \text{for all } v_j \in V\,, \tag{2.2}$$

$$y_e \geqslant 0 \quad \text{for all } e \in E\,. \tag{2.3}$$

The algorithm to be given maintains an integral primal feasible solution—(1.2) and (2.1) are satisfied and $x_j \in \{0, 1\}$ for all $v_j \in V$—and a complementary dual solution:

$$\text{if } y_e \neq 0 \text{ where } e = (v_i, v_j)\,, \text{then } x_i + x_j = 1\,, \tag{2.4}$$

$$\text{if } x_j \neq 0\,, \text{then } \sum_{e:\, v_j\in e} y_e = w_j\,. \tag{2.5}$$

Optimality is achieved when the dual feasibility conditions (2.2) and (2.3) are satisfied.

Without loss of generality, we assume that G is connected and contains at least two vertices. We call v_j a *black vertex* if $v_j \in P$ ($x_j = 1$) and call v_j a *white vertex* if $v_j \in V\backslash P = \bar{P}$ ($x_j = 0$).

To describe the primal and dual solutions graphically, we associate with an integral primal feasible solution a *spanning forest.* Each tree $T_r = (V_r, E_r)$ in the forest has a special white vertex v_r called the *root of* T_r. Each tree is *alternating* in the sense that all tree edges connect a white vertex and black vertex. The *cost* of a tree T_r, $c(T_r)$, is the sum of the black vertex weights of the tree minus the sum of the white vertex weights of the tree.

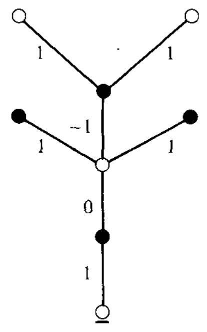
By deleting an edge e from a tree T_r we would obtain two trees T'_r and T'_s, where T'_r is the tree that contains v_r and v_s is the vertex of T'_s that is contained in e. T'_s is called the *branch of* e and is denoted by B_e; v_s is called the *root of* B_e. B_e is called a *white* or *black branch* according to whether its root is white or black. The cost of B_e, $c(B_e)$, is the cost of T'_s.

The *dual solution* corresponding to a given packing and associated spanning forest is defined by

$$y_e = \begin{cases} 0 & \text{if } e \text{ is not a tree edge}\,, \\ c(B_e) & \text{if } e \text{ is a tree edge and its branch is black}\,, \\ -c(B_e) & \text{if } e \text{ is a tree edge and its branch is white}\,. \end{cases} \tag{2.6}$$

Fig. 1 shows an alternating tree T_r rooted at v_r. The numbers on the edges are the y_e's.

Some important facts about the dual solution are summarized in Proposition 1.



$c(T_r) = 0$

Fig. 1. All vertex weights are equal to one. The underlined vertex is the root.

Proposition 1. *Given a packing P, associated spanning forest and dual solution* (2.6), *then*

(a) *if* v_j *is not a root of a tree, then the dual constraint* (2.2) *is an equality*;

(b) *the primal and dual solutions satisfy the complementary slackness conditions* (2.4) *and* (2.5);

(c) *dual infeasibilities occur only when*

(i) *a tree* T_r *has* $c(T_r) < 0$, *in which case* $\sum_{e:\, v_r \in e} y_e = w_r + c(T_r)$,

(ii) *a black branch* B_e *has* $c(B_e) < 0$, *in which case* $y_e = c(B_e)$,

(iii) *a white branch* B_e *has* $c(B_e) > 0$, *in which case* $y_e = -c(B_e)$.

Proof. (a) Let $e_0, e_1, \ldots, e_p$ be the edges adjacent to v_j with e_0 the edge on the unique path from v_j to the root. Since the tree is alternating, we have $y_{e_0} = w_j - \sum_{j=1}^{p} y_{e_j}$.

(b) Since roots are white vertices, (2.5) follows from (a); (2.4) is implied by $y_e = 0$ for non-tree edges and the alternating structure of the trees.

(c) Statements (ii) and (iii) follow directly from the definition of the y_e's; (i) is implied by $c(T_r) = -w_r + \sum_{e:\, v_r \in e} y_e$. □

If $c(T_r) \geqslant 0$ for each of the trees in the spanning forest and $y_e \geqslant 0$ for all $e \in E$, then the corresponding primal and dual solutions are optimal. However, if either of these conditions fails to hold, then the structure of the spanning forest and, perhaps, the packing are changed as described in the next section.

3. Fundamental algorithmic operations

The fundamental steps are classified in two ways. The first category deals with the packing and the second with the structure of the spanning forest and the dual solution.

A change in the packing P is called an *augmentation*. Let Q be the set of vertices in a tree or black branch. An augmentation switches the colors of all of the vertices in Q so that the new primal solution is given by $P' = P \cup (Q \cap \bar{P}) \backslash (Q \cap P)$. In order for P' to be a packing, it is necessary and sufficient that no edges exist joining a vertex in $Q \cap \bar{P}$ to a vertex in $(V \backslash Q) \cap P$. The necessity of this condition is obvious; sufficiency follows because, in a bipartite graph, no pair of white vertices in the same tree can be joined by an edge. Furthermore,

$$w(P') - w(P) = \begin{cases} -c(T_r) & \text{for a tree}\,, \\ -y_e & \text{for a black branch } B_e\,. \end{cases}$$

Thus, if there exists an *augmenting tree*—a tree T_r with $c(T_r) < 0$ and none of the white vertices in T_r is adjacent to a black vertex not in T_r, or an *augmenting branch*—a black branch B_e with $y_e < 0$ and none of the white vertices in B_e is adjacent to a black vertex not in B_e, then P' is a packing with $w(P') > w(P)$. If there is no augmenting tree or branch, then the algorithm will do a degenerate operation that does not change the packing.

There are six types of changes in the structure of the spanning forest as explained below and illustrated in Fig. 2. The first three follow augmentations and the last three are degenerate operations. It is obvious from the nature of these constructions that all of the operations preserve a spanning forest consisting of alternating trees with white roots and the dual variables satisfy the conditions (2.4) and (2.5).

Operation 1. *Changing a root.* Following an augmentation on a tree T_r with more than one vertex and $c(T_r) < 0$, $v_p \in V_r$ adjacent to v_r is selected as the new root to form the tree T'_p. Let $e' = (v_r, v_p)$. The dual variable adjustment is

$$y_{e'} \leftarrow y_{e'} - c(T_r)\,,$$

$$C(T'_p) \leftarrow -c(T_r)\,.$$

Given a tree T_q containing v_s, let $P_{q,s} = \{e\colon e \text{ a tree edge on the path joining } v_q \text{ to } v_s\}$, $P^{\mathrm{b}}_{q,s} = \{e \in P_{q,s}\colon B_e \text{ black}\}$ and $P^{\mathrm{w}}_{q,s} = P_{q,s} \backslash P^{\mathrm{b}}_{q,s}$.

Operation 2. *Attaching a black vertex to a tree.* Following an augmentation on a tree $T_r = \{v_r\}$, the tree T'_s is formed by attaching the edge $e' = (v_r, v_p)$ to

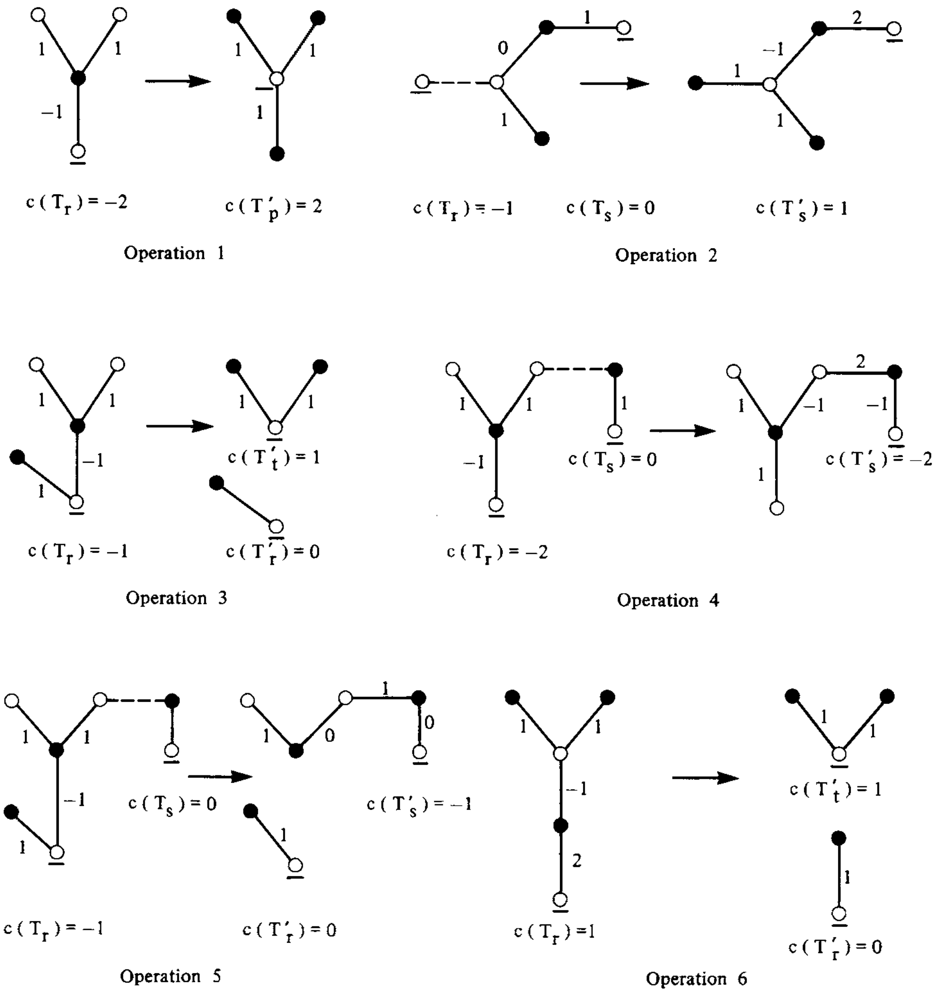


Fig. 2.

T_s, where $v_p \in V_s$ is a white vertex. The dual variable adjustment is

$$y_e \leftarrow y_e + w_r, \quad \forall e \in P^{\mathrm{b}}_{s,r},$$

$$y_e \leftarrow y_e - w_r, \quad \forall e \in P^{\mathrm{w}}_{s,r},$$

$$c(T'_s) \leftarrow c(T_s) + w_r.$$

Operation 3. *Cutting a black branch.* Following an augmentation on a black branch $B_{e'}$ of T_r with $y_{e'} < 0$, the edge $e' = (v_q, v_t)$ is deleted from T_r to obtain the trees T'_r and T'_t. The dual variable adjustment is

$$y_e \leftarrow y_e - y_{e'}, \quad \forall e \in P^{\mathrm{b}}_{r,q},$$

$$y_e \leftarrow y_e + y_{e'}, \quad \forall e \in P^{\mathrm{w}}_{r,q},$$

$$c(T'_r) \leftarrow c(T_r) - y_{e'},$$

$$c(T'_t) \leftarrow -y_{e'},$$

$$y_{e'} \leftarrow 0.$$

Operation 4. *Attaching one tree to another.* Let $v_p \in V_r$ be a white vertex, $c(T_r) < 0$, $v_h \in V_s$ a black vertex, $s \neq r$, and $\hat{e} = (v_p, v_h)$. T_r is attached to T_s to form T'_s by making $\hat{e}$ a tree edge. The dual variable adjustment is

$$y_e \leftarrow y_e + c(T_r), \quad \forall e \in P^{\mathrm{b}}_{s,r},$$

$$y_e \leftarrow y_e - c(T_r), \quad \forall e \in P^{\mathrm{w}}_{s,r},$$

$$c(T'_s) \leftarrow c(T_s) + c(T_r).$$

Operation 5. *Cutting a branch and attaching it to a tree.* As in Operation 3, a black branch $B_{e'}$, $e' = (v_q, v_t)$, with $y_{e'} < 0$ is cut from T_r. Then, as in Operation 4, $B_{e'}$ is attached to T_s to form the tree T'_s (here s and r may be the same) using the edge $\hat{e} = (v_p, v_h)$, where v_p is a white vertex of $B_{e'}$ and $v_h \in V_s$ is a black vertex. The dual variable adjustment is

$$y_e \leftarrow y_e - y_{e'}, \quad \forall e \in P^{\mathrm{b}}_{r,q} \cup P^{\mathrm{w}}_{s,t},$$

$$y_e \leftarrow y_e + y_{e'}, \quad \forall e \in P^{\mathrm{w}}_{r,q} \cup P^{\mathrm{b}}_{s,t},$$

$$c(T'_r) \leftarrow c(T_r) - y_{e'},$$

$$c(T'_s) \leftarrow c(T_s) + y_{e'},$$

$$y_{e'} \leftarrow 0.$$

(Note that if $s = r$, then $c(T'_r) = c(T_r)$, and if there exists $e \in P_{r,q} \cap P_{s,h}$, y_e is not changed since it is both increased and decreased by $y_{e'}$.)

Operation 6. *Cutting a white branch.* Given a white branch $B_{e'}$ of T_r, the edge $e' = (v_q, v_t)$ is deleted from T_r to obtain the trees T'_r and T'_t. The dual variable adjustment is

$$y_e \leftarrow y_e + y_{e'}, \quad \forall e \in P^{\mathrm{b}}_{r,q},$$

$$y_e \leftarrow y_e - y_{e'}, \quad \forall e \in P^{\mathrm{w}}_{r,q},$$

$$c(T'_r) \leftarrow c(T_r) + y_e,$$

$$c(T'_t) \leftarrow -y_e,$$

$$y_{e'} \leftarrow 0.$$

(Note that if $y_{e'} = 0$, there is no dual variable change.)

Identifying appropriate degenerate operations is the key to the finiteness of the algorithm. This requires a partition of the trees as described in the next section.

4. Tree partition and cutting of white branches

Each tree is partitioned into subtrees which are then assigned to one of three classes:

(1) the *deficient class* $\mathscr{D}$—each member is a subtree such that the sum of the weights of the white vertices is greater than the sum of the weights of the black vertices.

(2) the *neutral class* $\mathscr{N}$—each member is a subtree such that the sums of the weights of the white and black vertices are equal.

(3) the *saturated class* $\mathscr{S}$—each member is a subtree such that the sum of the weights of the white vertices is less than the sum of the weights of the black vertices.

Given a tree T_r, the partitioning, cutting, and determination of subtrees is done recursively. When a white branch is cut from T_r, the dual variables are adjusted in the remainder of the tree (see Operation 6) and then the two resulting trees are handled separately. When a black branch is placed in one of the classes, a temporary dual variable adjustment is made in the remainder of the tree as if the black branch had been cut (see Operation 3). After the partitioning, dual variables that have been temporarily adjusted are restored to their original values.

Let $E_r^{\rm w} = \{e \in E_r\colon B_e$ is white$\}$, $E_r^{\rm b} = \{e \in E_r\colon B_e$ is black$\}$, $E_r^{\rm w^-} = \{e \in E_r^{\rm w}\colon y_e < 0\}$, $E_r^{\rm w^0} = \{e \in E_r^{\rm w}\colon y_e = 0\}$, $E_r^{\rm b^-} = \{e \in E_r^{\rm b}\colon y_e < 0\}$ and $E_r^{b^0} = \{e \in E_r^{\rm b}\colon y_e = 0\}$.

Algorithm: *tree partition and cutting of white branches*

Given T_r, $\mathscr{N}$, $\mathscr{D}$ and $\mathscr{S}$:

If $E_r^{\rm w-} \cup E_r^{\rm b-} = \emptyset$

  then if $c(T_r) = 0$ **then** $\mathscr{N} \leftarrow \mathscr{N} \cup T_r$

    else if $c(T_r) < 0$

      then if $E_r^{\rm w^0} = \emptyset$ **then** $\mathscr{D} \leftarrow \mathscr{D} \cup T_r$

        else let $B_{e_1}, \ldots, B_{e_p}$ be the maximal branches of T_r with $e_i \in E_r^{\rm w^0}$. Cut these branches from T_r to form the trees T'_r and T'_{s_i}, $i = 1, \ldots, p$ (Operation 6)

        $\mathscr{D} \leftarrow \mathscr{D} \cup T'_r$ and $\mathscr{N} \leftarrow \mathscr{N} \bigcup_{i=1}^{p} T_{s_i}$

      else if $E_r^{\rm b^0} = \emptyset$ **then** $\mathscr{S} \leftarrow \mathscr{S} \cup T_r$

        else let $B_{e_1}, \ldots, B_{e_q}$ be the maximal branches of T_r with $e_i \in E_r^{\rm b^0}$.

        $\mathscr{N} \leftarrow \mathscr{N} \bigcup_{i=1}^{q} B_{e_i}$ and $\mathscr{S} \leftarrow \mathscr{S} \cup T'_r$

else if there exists $e' \in E_r^{b^-}$ with $y_e \geq 0 \; \forall \; e$ of $B_{e'}$ where v_s is the root of $B_{e'}$, **then** treat $B_{e'}$ as if it were a tree T_s' with $c(T_s') = y_e$ (but do not cut e' from T_r). Temporarily adjust dual variables in the remainder of the tree T_r' and then check T_r' and T_s' recursively.

else there exists $e' \in E_r^{w^-}$ with $y_e \geq 0 \; \forall \; e$ of $B_{e'}$ where v_s is the root of $B_{e'}$. Cut B_e from T_r to form the trees T_r' and T_s' (Operation 6). Adjust dual variables in T_r' and then check T_r' and T_s' recursively.

While there exists $B_{e'}, B_{\hat{e}} \in \mathscr{D}$ in the same tree, where $\hat{e}$ has one end in $B_{e'}$ and the other in $B_{\hat{e}}$, combine the two branches into the single branch $B_{e'}'$. $\mathscr{D} \leftarrow (\mathscr{D} \backslash (B_{e'} \cup B_{\hat{e}})) \cup B_{e'}'$.

repeat.

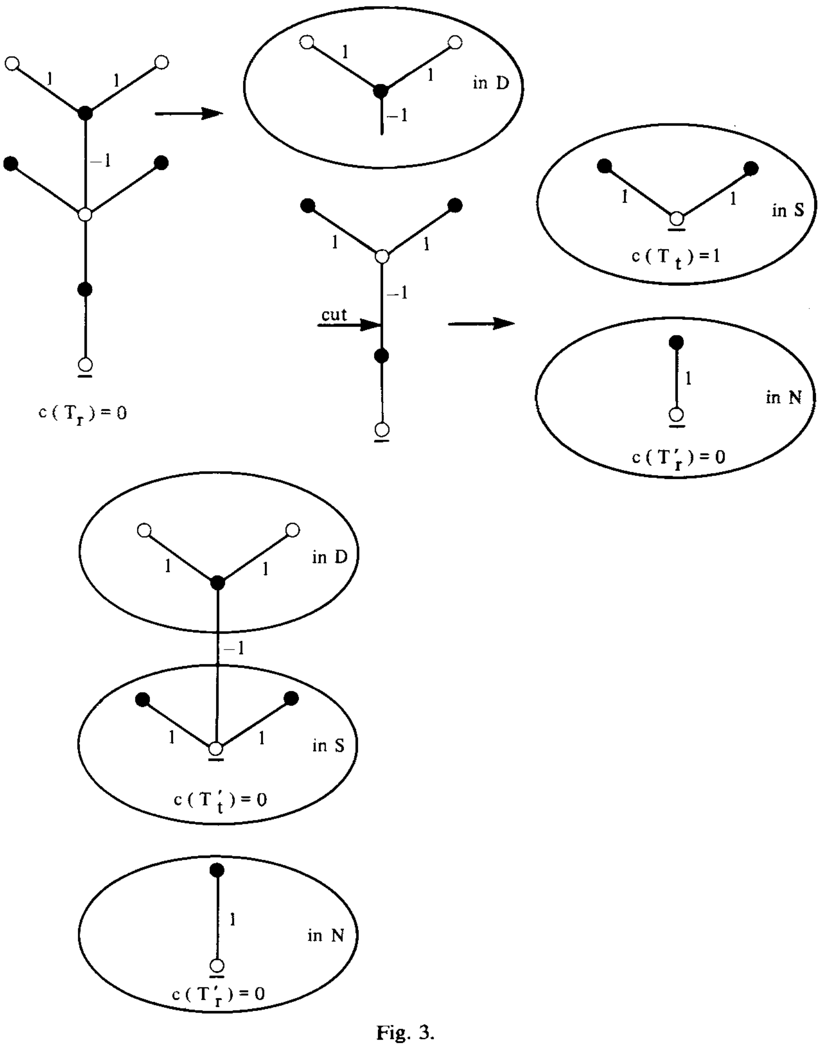


Fig. 3.

The algorithm is illustrated in Fig. 3. The output is a collection of partitioned trees $\{T_{r_1}, \ldots, T_{r_k}\}$.

Later we will need some properties of the partitioned trees that are simple consequences of the construction. These are given below.

(a) For $i = 1, \ldots, k$, $E_{r_i}^{w^-} = \emptyset$.

(b) A branch B_e of T_r which is a member of $\mathscr{D}$ does not contain any branches B_{e_i} with $e_i \in E_r^{w^0}$.

(c) $D_i \in \mathscr{D}$ implies D_i is a tree T_r with $c(T_r) < 0$ or a branch B_e of T_r with $e \in E_r^{b^-}$. Conversely, if $c(T_r) < 0$ or T_r contains a branch B_e with $e \in E_r^{b^-}$, then there exists $D_i \in \mathscr{D}$ such that D_i is included in T_r.

(d) If $T_r = D_i \in \mathscr{D}$, then $c(T_r) < 0$ and $c(T_r) - y_e < 0 \ \forall \ e \in E_r^b$.

(e) If $T_r = D_i \in \mathscr{D}$, then for $e_1 = (v_t, v_q) \in P_{r,p}^b$ and $\forall \ e \in P_{r,t}^w$, $y_e + y_{e_1} > 0$.

(f) Let v_p be a leaf of T_r. $P_{r,p}$ contains at most one member of $\mathscr{D}$, one member of $\mathscr{N}$ and one member of $\mathscr{S}$. If $P_{r,p}$ contains a member of $\mathscr{D}$, that member contains v_p. If $P_{r,p}$ contains a member of $\mathscr{S}$, that member contains v_r (see Fig. 4).

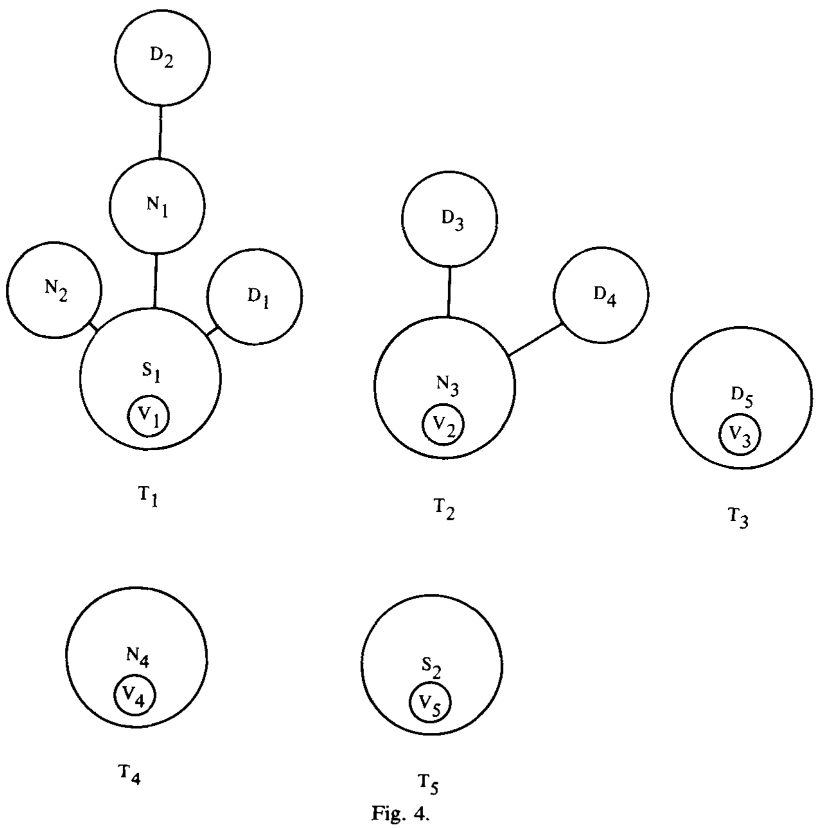


Fig. 4.

(g) If $T_r = S_i \in \mathcal{S}$, then $c(T_r) > 0$ and $y_e > 0 \; \forall \; e \in E_r^b$.
(h) If $T_r = N_i \in \mathcal{N}$, then $c(T_r) = 0$ and $y_e \geqslant 0 \; \forall \; e \in E_r^b$.

5. The algorithm

The essential idea of the algorithm is to achieve optimality by making $\mathcal{D} = \emptyset$.

Step 0 (*initialization*)

Let $P \leftarrow \emptyset$, $T_r \leftarrow \{v_r\}$, $c(T_r) \leftarrow -w_r$, $\forall \; v_r \in V$, $y_e \leftarrow 0 \; \forall \; e \in E$, $\mathcal{S} \leftarrow \emptyset$, $\mathcal{N} \leftarrow \emptyset$, $\mathcal{D} = \{\{v_i\}: v_i \in V\}$.

Step 1 (*checking optimality*)

If $\mathcal{D} = \emptyset$
  then output P and $\{y_e: e \in E\}$ as the optimal primal and dual solutions.
  else if there exists $D_i \in \mathcal{D}$ that is an augmenting tree T_r
    then go to Step 2.
    else if There exists $D_i \in \mathcal{D}$ that is an augmenting branch $B_{e'}$
      then go to Step 3.
      else go to Step 4.

Step 2 (*augmentation on a tree*)

If $T_r = \{v_r\}$
  then $P \leftarrow P \cup \{v_r\}$. Execute Operation 2 (attaching a black vertex to a tree).
    $T_r \leftarrow \emptyset$, $T_s \leftarrow T'_s$, $h \leftarrow s$.
    Go to Step 5.
  else $P \leftarrow P \cup (V_r \cap \bar{P}) \backslash (V_r \cap P)$. Execute Operation 1 (changing a root).
    $T_r \leftarrow \emptyset$, $T_p \leftarrow T'_p$, $h \leftarrow p$.
    Go to Step 5.

Step 3 (*augmentation on a branch*)

Let Q be the vertices of $B_{e'}$. $P \leftarrow P \cup (Q \cap \bar{P}) \backslash (Q \cap P)$. Execute Operation 3 (cutting a black branch). $T_r \leftarrow T'_r$, $T_t \leftarrow T'_t$, $h \leftarrow t$.

Go to Step 5.

Step 4 (*attaching a deficient member to a tree*)

There exists a non-tree edge $\hat{e} = (v_p, v_h)$, where $v_p \in \bar{P}$ belongs to $D_i \in \mathcal{D}$ and $v_h \in P$ belongs to T_s.

If D_i is a tree T_r
  then execute Operation 4 (attaching one tree to another).
    $T_r \leftarrow \emptyset$, $T_s \leftarrow T'_s$, $h \leftarrow s$.
    Go to Step 5.
  else Let D_i be the branch $B_{e'}$ of T_r. Execute Operation 5 (cutting a branch and attaching it to a tree). $T_r \leftarrow T'_r$, $T_s \leftarrow T'_s$, $h \leftarrow s$.
    Go to Step 5.

Step 5 (*tree partition and cutting of white branches*)

Apply this algorithm (see the previous section) to T_h.
Go to Step 1.

6. Complexity of the algorithm

When the algorithm stops ($\mathscr{D} = \emptyset$) in Step 1, properties (a) and (c) of the partitioned trees imply that dual feasibility has been achieved. Thus the primal and dual solutions are optimal.

Suppose $D_i \in \mathscr{D}$. Define the *deficiency of* D_i:

$$d(D_i) = \begin{cases} -c(T_r) > 0 & \text{if } D_i \text{ is a tree } T_r, \\ -y_e > 0 & \text{if } D_i \text{ is a black branch } B_e, \end{cases}$$

and the *deficiency of* $\mathscr{D}$ to be $d(\mathscr{D}) = \Sigma_{D_i \in \mathscr{D}} d(D_i)$. If the optimality conditions are not satisfied, $d(\mathscr{D}) > 0$.

We now investigate how $d(\mathscr{D})$ changes between successive calls of Step 1, which we call an *iteration*. Suppose $(\mathscr{D} \neq \emptyset, \hat{\mathscr{D}})$, $(\{y_e\}, \{\hat{y}_e\})$, $(\{c(T_r)\}, \{c\{\hat{T}_r)\})$ and $(P, \hat{P})$ are the deficient class, dual variables, tree costs and packing before and after some iteration, respectively.

Proposition 2. *If an iteration involves an augmentation,*

$$d(\mathscr{D}) - d(\hat{\mathscr{D}}) \geq w(\hat{P}) - w(P) > 0\,.$$

Proof. (a) Suppose the augmentation occurs on D_1, which is a tree T_r with more than one vertex (see Operation 1). We have $c(T'_p) = -c(T_r) > 0$ and $y'_{e'} = y_{e'} - c(T_r) \geq 0$ by property (a) of tree partitioning. Thus we have $E_p^{b^-} = \emptyset$ after Step 2 since $E_r^{w^-} = \emptyset$ by property (a). Now if $E_r^{b^-} \neq \emptyset$, then $E_p^{w^-} \neq \emptyset$ and it will be necessary to cut white branches from T'_p. However, each of these branches yields a tree whose pieces belong to either $\mathscr{N}$ or $\mathscr{S}$. Finally, after cutting such branches, the tree $\hat{T}_p$ that remains has $E_p^{b^-} = \emptyset$, for if there existed $e \in E_p^{b^-}$ in $\hat{T}_p$, then in a subtree of T_r during the partitioning algorithm, we would have had $e \in E_r^{w^-}$, which is not possible since such branches would have been cut. Thus the collection of trees $\{\hat{T}_{r_1}, \ldots, \hat{T}_{r_k}\}$ generated from T_r have all of their pieces contained in members of $\mathscr{S} \cup \mathscr{N}$, $\hat{\mathscr{D}} = \mathscr{D} \backslash \{D_1\}$ and $d(\mathscr{D}) - d(\hat{\mathscr{D}}) = w(\hat{P}) - w(P) = -c(T_r) > 0$.

(b) Suppose the augmentation occurs on D_1, which is a branch $B_{e'}$ of T_r (see Operation 3). The deficiency of the vertices in the subtree T'_r has not changed since T'_r was partitioned as if $B_{e'}$ had been cut. The argument that the deficiency of the vertices in $B_{e'}$ is reduced to zero is identical to the argument in

(a) of this proof. Hence $\hat{\mathscr{D}} = \mathscr{D}\backslash\{D_1\}$ and $d(\mathscr{D}) - d(\hat{\mathscr{D}}) = w(\hat{P}) - w(P) = -y_e > 0$.

(c) Suppose the augmentation occurs on D_1, which is a tree $T_r = \{v_r\}$ and v_r is attached to T_s by $e' = (v_r, v_p)$ to obtain T'_s (see Operation 2). Clearly, $D_1 \not\in \hat{\mathscr{D}}$. To show that the deficiency of the collection of vertices of T'_s following Step 5 is not larger than the deficiency of the vertices of T_s prior to Operation 2, it suffices to prove that $c(\hat{T}_s) \geq c(T_s)$ and $\hat{y}_e \geq y_e\ \forall\ e \in P^{\mathrm{b}}_{s,r}$.

After Operation 2 we have $y'_e - y_e = w_r > 0\ \forall\ e \in P^{\mathrm{b}}_{s,r}$ and $c(T'_s) - c(T_s) = w_r$. Thus, it suffices to show that if any white branches are cut in the partitioning procedure, $\hat{y}_e \geq y'_e - w_r\ \forall\ e \in P^{\mathrm{b}}_{s,r}$ and $c(\hat{T}_s) \geq c(T'_s) - w_r$. We have $y_e \geq 0\ \forall\ e \in P^{\mathrm{w}}_{s,p}$ in T_s, which implies $y'_e \geq -w_r\ \forall\ e \in P^{\mathrm{w}}_{s,p}$ in T'_s. The result now follows since the change of dual variables in the cutting of white branches implies $c(\hat{T}_s) \geq c(T'_s) + \min\{y'_e\colon e \in P^{\mathrm{w}}_{s,p}\}$ and $\forall\ e \in P^{\mathrm{b}}_{s,r}\ \hat{y}_e \geq y'_e + \min\{y'_e\colon e \in P^{\mathrm{w}}_{s,p}\}$. □

Proposition 3. *If an iteration involves a degenerate operation (Step 4) and a member of $\mathscr{D}$ is connected to*

(a) *a member of $\mathscr{S}$, then $d(\hat{\mathscr{D}}) > d(\mathscr{D})$ (deficient–saturated connection);*

(b) *a member of $\mathscr{N}$, then $d(\hat{\mathscr{D}}) = d(\mathscr{D})$, $|\hat{\mathscr{D}}| = |\mathscr{D}|$ and $|\{v \in \hat{D}_i\colon \hat{D}_i \in \hat{\mathscr{D}}\}| > |\{v \in D_i\colon D_i \in \mathscr{D}\}|$ (deficient–neutral connection);*

(c) *a member of $\mathscr{D}$, then $d(\hat{\mathscr{D}}) = d(\mathscr{D})$ and $|\hat{\mathscr{D}}| < |\mathscr{D}|$ (deficient–deficient connection).*

Proof. We assume that D_1 is joined to either D_2, N_1 or S_1, where the latter element belongs to the tree T_s, to form the tree T'_s. Without loss of generality, we can assume that D_1 is a tree T_r, since if it were a branch $B_{e'}$ of T_r, the deficiency of the vertices in the subtree T'_r would not change. Thus, we are dealing with Operation 4 and T_r is joined to T_s by the edge $\hat{e} = (v_p, v_h)$.

In T'_s, $P^{\mathrm{w}}_{s,r} = P^{\mathrm{w}}_{s,h} \cup P^{\mathrm{b}}_{r,p} \cup \{\hat{e}\}$ and $P^{\mathrm{b}}_{s,r} = P^{\mathrm{b}}_{s,h} \cup P^{\mathrm{w}}_{r,p}$. First we will show that no white branches are cut in the partitioning of T'_s. In T'_s, $\forall\ e \in P^{\mathrm{w}}_{s,h} \cup \{\hat{e}\}$ we have $y'_e = y_e - c(T_r) \geq -c(T_r) > 0$ since $y_e \geq 0$ by property (a) of partitioning; $\forall\ e \in P^{\mathrm{b}}_{r,p}$ we have $y'_e = y_e - c(T_r) > 0$ by property (d) of partitioning. Thus $y'_e > 0$ $\forall\ e \in P^{\mathrm{w}}_{s,r}$ so that if no temporary dual variable adjustments occur, no white branches will be cut. Now suppose that B_{e_1} is the first black branch of T'_s that yields a temporary dual variable adjustment. Consider $e_1 = (v_q, v_t) \in P^{\mathrm{b}}_{s,h}$ so that only dual variables on the path $P_{s,t}$ change. We have $\forall\ e \in P^{\mathrm{w}}_{s,t}$, $\tilde{y}_e = y'_e + y'_{e_1} = y_e + y_{e_1}$. Since $y_e \geq 0$, $\tilde{y}_e < 0$ implies $y_{e_1} < 0$. But $\tilde{y}_e < 0$ and $y_{e_1} < 0$ imply that B_e would have been cut when T_s was partitioned. Now suppose that $e_1 \in P^{\mathrm{w}}_{r,p}$; $\forall\ e \in P^{\mathrm{w}}_{s,h} \cup \{\hat{e}\}$ we have $\tilde{y}_e = y_e + y_{e_1} > 0$ since $y_e \geq 0$ by property (a) and $y_{e_1} > 0$ by property (b). Finally, $\forall\ e \in P^{\mathrm{w}}_{s,t} \cap P^{\mathrm{b}}_{r,p}$ we have $\tilde{y}_e = y_e + y_{e_1} > 0$ by property (e). The same argument applies to multiple, temporary dual variable changes so we have shown that no branches will be cut from T'_s in Step 5.

Suppose that v_h belongs to $S_1 \in \mathscr{S}$. By property (f), S_1 contains v_s. T_s may also contain members of $\mathscr{D} \cup \mathscr{N}$, but since these members would have no effect on the change in deficiency, it suffices to assume that $T_s = S_1$, which is notationally easier to treat. To show that the deficiency of the collection of vertices in $D_1 \cup S_1$ is less than $-c(T_r)$, it is sufficient to show that $y'_e > c(T_r)\ \forall\, e \in P^{\mathrm{b}}_{s,r}$ and $c(T'_s) > c(T_r)$. We have $y'_e = y_e + c(T_r) > c(T_r)\ \forall\, e \in P^{\mathrm{w}}_{r,p}$ since $y_e > 0$ by properties (a) and (b). We have $y'_e = y_e + c(T_r) > c(T_r)\ \forall\, e \in P^{\mathrm{b}}_{s,h}$ and $c(T'_s) = c(T_s) + c(T_r) > c(T_r)$ since $y_e > 0$ and $c(T_s) > 0$ by property (g). Since $P^{\mathrm{b}}_{s,r} = P^{\mathrm{w}}_{r,p} \cup P^{\mathrm{b}}_{s,h}$, the proof of (a) is complete.

Now suppose that v_h belongs to $N_1 \in \mathscr{N}$. T_s may contain members of $\mathscr{D}$ attached to N_1 and a member of $\mathscr{S}$ attached to N_1 as explained in property (f) and illustrated in Fig. 4. However, these members would have no effect on the change in deficiency so it suffices to treat the notationally simpler case $T_s = N_1$. To show that the deficiency of the collection of vertices in $D_1 \cup N_1$ equals $-c(T_r)$, it is sufficient to show that $y'_e \geqslant c(T_r)\ \forall\, e \in P^{\mathrm{b}}_{s,r}$ and $c(T'_s) = c(T_r)$. We have $y'_e > c(T_r)\ \forall\, e \in P^{\mathrm{w}}_{r,p}$ as explained above. We have $y'_e = y_e + c(T_r) \geqslant c(T_r)$ $\forall\, e \in P^{\mathrm{b}}_{s,h}$ and $c(T'_s) = c(T_s) + c(T_r) = c(T_r)$ since $y_e \geqslant 0$ and $c(T_s) = 0$ by property (h). Furthermore, $\min\{y'_e\colon e \in P^{\mathrm{w}}_{r,p}\} > c(T_r) = c(T'_s)$ which implies that a member $\hat{D}_1$ of $\hat{\mathscr{D}}$ will be formed that contains D_1 and at least the edge $\hat{e} = (v_p, v_h)$; the remainder of N_1 will be put in $\mathscr{N}$. Thus $\hat{D}_1$ has the same deficiency as D_1, but a larger number of vertices, which completes the proof of (b).

Finally, suppose that v_h belongs to $\mathscr{D}_2 \in \mathscr{D}$. As in the two previous cases, other members of T_s would have no effect on the change in deficiency so we treat the case in which $T_s = D_2$. To show that the deficiency of the collection of vertices in $D_1 \cup D_2$ equals $-c(T_r) - c(T_s)$, it is sufficient to show that $y'_e > c(T_r) + c(T_s)\ \forall\, e \in P^{\mathrm{b}}_{s,r}$ and $c(T'_s) = c(T_r) + c(T_s)$. We have $y'_e > c(T_r)\ \forall\, e \in P^{\mathrm{w}}_{r,p}$ as explained above, which implies $y'_e > c(T_r) + c(T_s)$ since $c(T_s) < 0$. We have $y'_e = y_e + c(T_r) > c(T_r) + c(T_s)\ \ \forall\, e \in P^{\mathrm{b}}_{s,h}$ since $y_e > c(T_s)$. Since $c(T'_s) = c(T_r) + c(T_s) < \min\{y'_e\colon e \in P^{\mathrm{b}}_{s,r}\}$, $\hat{D}_1 = D_1 \cup D_2 \cup \{\hat{e}\}$ becomes a member of $\hat{\mathscr{D}}$ with deficiency equal to $-c(T_r) - c(T_s)$, which proves (c). □

Assume that w_r is an integer for all $v_r \in V$. Let $M = \Sigma_{r:\, v_r \in V}\, w_r$ and $n = |V|$.

Theorem 1. *The algorithm requires at most $(n-1)M$ iterations and its running time is $\mathrm{O}(n^3 M)$.*

Proof. Initially $d(\mathscr{D}) = M$ and each augmentation and deficient-saturated connection reduces $d(\mathscr{D})$ by a positive integer. Therefore, the number of these iterations cannot exceed M. Between these iterations there can be a sequence of deficient–neutral and deficient–deficient connections. Suppose following a reduction in $d(\mathscr{D})$, we have $|\mathscr{D}| = k_1$, and the members of $\mathscr{D}$ collectively contain k_2 vertices, $k_2 \geqslant k_1$. Then from Proposition 3(b) and (c), prior to the next decrease in $d(\cdot)$, there can be at most $n - k_2$ deficient–neutral connections and $k_1 - 1$ deficient–deficient connections. Consequently, the length of such a sequence is bounded by $k_1 - 1 + n - k_2 < n$, and the total number of iterations is bounded by $(n-1)M$.

Each dual change involves $O(n)$ calculations to change the dual variables along some path, and the number of dual changes (including temporary ones) per iteration is also $O(n)$ since there can be one for each tree edge. The remaining work of an iteration, e.g., changing the packing is also $O(n)$. Hence the running time per iteration is $O(n^2)$ and the total running time is $O(n^3M)$. □

Observe that Theorem 1 still applies if we modified the algorithm by removing the priority on augmentations. In other words, it is not necessary to search for the existence of augmentations. So long as $\mathcal{D} \neq \emptyset$, a completely arbitrary choice of one of its members suffices in Step 1 to obtain the result of Theorem 1.

7. A scaling method for weighted vertex packing problems

Here we use a scaling method developed by Edmonds and Karp [5] for transportation problems and show that by a recursive adjustment of the weights the running time of our algorithm becomes polynomial in the problem input.

Define l by

$$2^l \leq \max_j w_j < 2^{l+1} \tag{7.1}$$

and change the weight w_j to $w_j^0 = \lfloor w_j/2^l \rfloor$, for all j. From (7.1) we have $w_j^0 \in \{0, 1\}$. The weighted vertex packing problem on G with weights $\{w_j^0\}$ is a cardinality vertex packing problem on the subgraph $G^0 = (V^0, E^0)$, where $V^0 = \{v_j \mid w_j^0 = 1\}$. Hence, we can solve this problem with our simplex method in at most n^2 steps. Let P^0 be the optimal packing and F^0 the spanning forest obtained at optimality.

For $i = 1, 2, \ldots, l$, suppose we have a packing P^{i-1} and a spanning forest F^{i-1} that represent an optimal solution to the weighted vertex packing problem with weights $\{w_j^{i-1}\}$. Then define new weights by $w_j^i = \lfloor w_j/2^{l-i} \rfloor$, for all j. Now we solve the new problem with weights w_j^i for all j. The packing P^{i-1} and the forest F^{i-1} form a feasible basis for this problem, since only the objective function has been changed.

Proposition 4. *Let d_i be the deficiency of the optimal solution (P^{i-1}, F^{i-1}) with respect to the problem with weights w_j^i for all j. Then $d_i < n$, $i = 1, 2, \ldots, l$.*

Proof. For all j, $2w_j^{i-1} \leqslant w_j^i \leqslant 2w_j^{i-1} + 1$. With the weight $2w_j^{i-1}$ on v_j for all j, (P^{i-1}, F^{i-1}) is an optimal solution; hence the deficiency is zero. Let $D_1, \ldots, D_q$ be the members of $\mathcal{D}$ obtained from the new weights. We will show that $d(D_j) \leqslant |D_j \cap \bar{P}^{i-1}|$ for all j, where $\bar{P}^{i-1} = V \backslash P^{i-1}$. Suppose with respect to the forest F^{i-1}, D_j is in a black branch B_e. Let $B_{e_1}, \ldots, B_{e_r}$ be the white branches in B_e such that $D_j = B_e \backslash \bigcup_{k=1}^r B_{e_k}$. Note that r may be zero, i.e., $D_j = B_e$.

We write for a subset $S \subseteq V$, $w^i(S) = \sum_{j:\, v_j \in S} w_j^i$ and $w^{i-1}(S) = \sum_{j:\, v_j \in S} w_j^{i-1}$. Then

$$
\begin{aligned}
d(D_j) &= w^i(D_j \cap \bar{P}^{i-1}) - w^i(D_j \cap P^{i-1}) \\
&\leqslant 2\{w^{i-1}(D_j \cap \bar{P}^{i-1}) - w^{i-1}(D_j \cap P^{i-1})\} + |D_j \cap \bar{P}^{i-1}| \,. \qquad (7.2)
\end{aligned}
$$

With the weights w_j^{i-1}, the white branch B_{e_k} was not cut; hence,

$$
0 \leqslant y_{e_k} = w^{i-1}(B_{e_k} \cap \bar{P}^{i-1}) - w^{i-1}(B_{e_k} \cap P^{i-1}) \quad \text{for all } k\,. \qquad (7.3)
$$

Combining (7.2) and (7.3), we get

$$
\begin{aligned}
d(D_j) &\leqslant 2\Big\{ w^{i-1}(D_j \cap \bar{P}^{i-1}) + \sum_{k=1}^{r} w^{i-1}(B_{e_k} \cap \bar{P}^{i-1}) - w^{i-1}(D_j \cap P^{i-1}) \\
&\quad - \sum_{k=1}^{r} w^{i-1}(B_{e_k} \cap P^{i-1}) \Big\} + |D_j \cap \bar{P}^{i-1}| \\
&= 2\{w^{i-1}(B_e \cap \bar{P}^{i-1}) - w^{i-1}(B_e \cap P^{i-1})\} + |D_j \cap \bar{P}^{i-1}| \\
&= 2(-y_e) + |D_j \cap \bar{P}^{i-1}|
\end{aligned}
$$

where y_e is the value of the dual variable associated with B_e in the optimal solution (P^{i-1}, F^{i-1}). Since $y_e \geqslant 0$, $d(D_j) \leqslant |D_j \cap \bar{P}^{i-1}|$.

Therefore, the total deficiency d_i is given by

$$
d_i = \sum_{j=1}^{q} d(D_j) \leqslant \sum_{j=1}^{q} |D_j \cap \bar{P}^{i-1}| < n\,.
$$

The same argument applies if D_j contains a root of a tree. □

Since, for $i = 1, \ldots, l$, the initial deficiency in the ith problem is bounded by n, the number of iterations for the ith problem is bounded by n^2. Therefore, we have the following theorem.

Theorem 2. *By the recursive scaling method, the algorithm finds an optimal solution in at most $n^2(l+1)$ iterations.*

Of course, it would be nice to have a primal simplex algorithm whose running time is independent of the vertex weights. We expect that either the algorithm given or an appropriate modification of it enjoys this property and hope to obtain such a result.

8. Relation to the primal simplex method

The algorithm that has been given is a primal 'simplex-like' method. It maintains a basic feasible solution to the linear probram

$$\max wx$$
$$Ax + Is = \mathbf{1}\,, \quad x, s \geqslant 0$$

where $A = (a_1, \ldots, a_n)$ is the $m \times n$ edge-vertex incidence matrix of G, $I = (h_1, \ldots, h_m)$ is an $m \times m$ identity matrix, 1 is a vector of ones, and $w = (w_1, \ldots, w_n)$. x_j is the variable for v_j, $j = 1, \ldots, n$, and s_i is the slack variable for e_i, $i = 1, \ldots, m$.

Each operation given in Section 3 corresponds to a primal simplex pivot in the sense that the pivot element in the simplex tableau is positive. We will describe this result without all of the details; these can be found in [10].

We characterize a basic feasible solution by the basis matrix $B = (b_1, \ldots, b_m)$ where, without loss of generality, we assume that $b_i = a_i$, $i = 1, \ldots, k$, $b_i = h_i$, $i = k+1, \ldots, m$. Thus the set of vertices whose variables are basic and non-basic, respectively, are $V^{\mathrm{B}} = \{v_1, \ldots, v_k\}$ and $V^{\mathrm{N}} = \{v_{k+1}, \ldots, v_n\}$, and the set of edges whose slack variables are basic and non-basic, respectively, are $E^{\mathrm{B}} = \{e_{k+1}, \ldots, e_m\}$ and $E^{\mathrm{N}} = \{e_1, \ldots, e_k\}$.

Proposition 5. *The graph $F = (V, E^{\mathrm{N}})$ is a spanning forest of G, with the collection of trees $\{T_j\colon v_j \in V^{\mathrm{N}}\}$. T_j contains $v_j \in V^{\mathrm{N}}$ and is alternating in the sense that each of its edges joins a white vertex to a black vertex.*

For $v_r \in V^{\mathrm{N}}$ let $\bar{a}_r = B^{-1}a_r$ and for $e_j \in E^{\mathrm{N}}$ let $\bar{h}_j = B^{-1}h_j$. Since B is totally unimodular, all of the non-zero components of $\bar{a}_r$ and $\bar{h}_j$ equal ± 1. Let $\bar{a}_r = (\bar{a}_{1r}, \ldots, \bar{a}_{mr})^{\mathrm{T}}$ and $\bar{h}_j = (\bar{h}_{1r}, \ldots, \bar{h}_{mj})^{\mathrm{T}}$.

Proposition 6. (a) *For $i = 1, \ldots, k$, $\bar{a}_{ir} = 1$ (-1) if and only if v_i is a black (white) vertex of T_r.*

(b) *For* $i = k+1, \ldots, m$, $\bar{a}_{ir} = 1$ (-1) *if and only if* $e_i \in E^{\mathrm{B}}$ *meets a single vertex of* T_r *which is white (black).*

(c) *For* $i = 1, \ldots, k$, $\bar{h}_{ij} = 1$ (-1) *if and only if* v_i *is in* B_{e_j}, *and* v_i *and* B_{e_j} *are of the same (opposite) color.*

(d) *For* $i = k+1, \ldots, m$, $\bar{h}_{ij} = 1$ (-1) *if and only if* $e_i \in E^{\mathrm{B}}$ *meets a single vertex of* B_{e_j} *and that vertex is of the opposite (same) color as* B_{e_j}.

Let $w^{\mathrm{B}} = (w_1, \ldots, w_k, 0_{k+1}, \ldots, 0_m)$, $\bar{w}_r = w_r - w^{\mathrm{B}}\bar{a}_r$ for $v_r \in V^{\mathrm{N}}$ and $z_j = -w^{\mathrm{B}}\bar{h}_j$ for $e_j \in E^{\mathrm{N}}$. These are the reduced costs in the simplex method. The following result is a direct consequence of Proposition 6.

Proposition 7. $\bar{w}_r = -c(T_r)$ *for all* $v_r \in V^{\mathrm{N}}$, $\bar{z}_j = -y_{e_j}$ *for all* $e_j \in E^{\mathrm{N}}$.

Propositions 6 and 7 provide the characteristics of the simplex tableau needed to describe the operations of Section 3 in linear programming terms. These are given in the following proposition.

Proposition 8. (a) *Operation* 1. $\bar{w}_r > 0$; $s_i = 1$ for all $i \in \{k+1, \ldots, m\}$ such that $\bar{a}_{ir} = 1$ (both ends of e_i are white vertices); there exists p, $1 \leqslant p \leqslant k$ such that $x_p = \bar{a}_{pr} = 1$ and $(v_p, v_r) \in E^{\mathrm{N}}$. The simplex method executes a non-degenerate pivot by making x_r basic and x_p non-basic.

(b) *Operation* 2. $\bar{w}_r = w_r > 0$; $\bar{a}_{ir} = 0$, $i = 1, \ldots, k$; $s_i = 1$ for all $i \in \{k+1, \ldots, m\}$ such that $\bar{a}_{ir} = 1$; there exists i^*, $k+1 \leqslant i^* \leqslant m$, such that $\bar{a}_{i^*r} = 1$. The simplex method executes a non-degenerate pivot by making x_r basic and s_{i^*} non-basic.

(c) *Operation* 3. $\bar{z}_j > 0$; B_{e_j} is a black branch with root v_t so that $\bar{a}_{tj} = x_t = 1$; $s_i = 1$ for all $i \in \{k+1, \ldots, m\}$ such that $\bar{a}_{ij} = 1$. The simplex method executes a non-degenerate pivot by making s_j basic and x_t non-basic.

(d) *Operation* 4. $\bar{w}_r > 0$; there exists i^*, $k+1 \leqslant i^* \leqslant m$, such that $\bar{a}_{i^*r} = 1$ and $s_{i^*} = 0$ (the end of e_{i^*} not in T_r is black). The simplex method does a degenerate pivot by making v_r basic and s_{i^*} non-basic.

(e) *Operation* 5. $\bar{z}_j > 0$; there exists i^*, $k+1 \leqslant i^* \leqslant m$, such that $\bar{a}_{i^*j} = 1$ and $s_{i^*} = 0$. The simplex method does a degenerate pivot by making s_j basic and s_{i^*} non-basic.

(f) *Operation* 6. B_{e_j} is a white branch with root v_t so that $\bar{a}_{tj} = 1$ and $x_t = 0$. The simplex method does a degenerate pivot by making s_j basic and x_t non-basic.

Operations 1–5 are done in Steps 2–4 of the algorithm. They correspond to choosing a non-basic variable with positive reduced cost and pivoting. Since, in Step 1, all white branches have non-positive reduced costs and, as noted previously, it is not necessary to do augmentations prior to degenerate operations, these are standard simplex steps. However in Step 5, the pivoting (Operation 6) is non-standard. Pivoting is only done on white branches, which must be distinguished, and these (because of the temporary dual variable adjustments) may have non-positive reduced costs.

We close by observing that our work can also be interpreted as a dual simplex method for (ELP), and thus can be considered as being complementary to recent developments on primal simplex algorithms for network flow problems by Cunningham [3, 4] and Barr, Glover and Klingman [1].

Acknowledgment

We are grateful to an anonymous referee, whose very careful reading and comments improved the clarity of this paper.

References

[1] R.S. Barr, F. Glover and D. Klingman, The alternative basis algorithm for assignment problems, Math. Programming 13 (1977) 1–13.

[2] V. Chvátal, On certain polytopes associated with graphs, J. Combin. Theory Ser. B 18 (1975) 138–154.

[3] W.H. Cunningham, A network simplex method, Math. Programming 11 (1976) 105–116.

[4] W.H. Cunningham, Theoretical properties of the network simplex method, Math. Oper. Res. 4(2) (1979) 196–208.

[5] J. Edmonds and R.M. Karp, Theoretical improvements in algorithmic efficiency for network flow problems, J. Assoc. Comput. Mach. 19(2) (1972) 248–264.

[6] D.R. Fulkerson, On the perfect graph theorem, in: T.C. Hu and S.M. Robinson, eds., Mathematical Programming (Academic Press, New York, 1973) 69–76.

[7] M.R. Garey, D.S. Johnson and L. Stockmeyer, Some simplified NP-complete graph problems, Theoret. Comput. Sci. 1 (1976) 237–267.

[8] M. Grötschel, L. Lovász and A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, Rept. No. 80151, 1980, University of Bonn. Combinatorica, to appear.

[9] W.-L. Hsu, Y. Ikura and G.L. Nemhauser, A polynomial algorithm for maximum weighted vertex packings on graphs without long odd cycles, Math. Programming 20 (1981) 225–232.

[10] Y. Ikura, Algorithms for vertex packing and clique covering problems on some unimodular graphs, Ph.D. Dissertation, Cornell University, 1981.

[11] L.G. Khachian, A polynomial algorithm in linear programming, Dokl. Akad. Nauk SSSR 244 (1979) 1093–1096; translated in Soviet Math. Dokl. 20 (1979) 191–194.

[12] L. Lovász, Normal hypergraphs and the perfect graph conjecture, Discrete Math. 2 (1972) 253–266.

[13] G.J. Minty, On maximal independent sets of vertices in claw-free graphs, J. Combin. Theory Ser. B 28 (1980) 284–304.
[14] S. Poljak, A note on stable sets and coloring of graphs, Comment. Math. Univ. Carolinae 15 (1974) 307–309.
[15] N. Sbihi, Algorithme de recherche d'un stable de cardinalité maximum dans un graphe sans étoile, Discrete Math. 29 (1980) 53–76.

Annals of Discrete Mathematics 16 (1982) 169–187

DEGREE-TWO INEQUALITIES, CLIQUE FACETS, AND BIPERFECT GRAPHS

Ellis L. JOHNSON*

IBM Research Center, Yorktown Heights, NY 10598, USA

Manfred W. PADBERG

Graduate School of Business Administration, New York University, New York, NY 10006, USA

We consider zero–one solutions to inequality systems having inequalities involving only two variables. Special cases include the independent node problem, the node covering problem, and the set packing problem. We develop a unified framework using a class of bidirected, transitively closed graphs and show properties of the zero–one solutions in terms of these graphs. A type of clique inequality is defined and shown to yield facets of the convex hull of zero–one solutions, extending well-known results for the independent node problem. For the latter problem, the development of perfect graphs has been important with regard to characterizing polytopes which have only zero–one vertices. We place our work in that setting, give some results, and conjecture some properties of perfect graphs involving our more general graph structures.

1. Introduction

We consider inequality systems with zero–one variables where the inequalities have only two variables per inequality. Without loss of generality, there are four possible inequalities for any pair x_i, x_j of variables:

$$\begin{aligned} x_i + x_j &\geq 1 , \\ x_i - x_j &\geq 0 , \\ -x_i + x_j &\geq 0 , \\ -x_i - x_j &\geq -1 . \end{aligned}$$

Our main object of study is the convex hull of zero–one solutions to inequality systems made up of inequalities of this type.

Besides being interesting as inequality systems in their own right, as we consider them here, such inequalities arise as logical implications of general

* Supported by Sonderforschungsbereich 21 (DFG), Institut für Operations Research, Universität Bonn, W. Germany.

zero–one programming problems and have been called degree-two constraints [1, 7, 8, 12]. For example, Guignard and Spielberg [7] build up logical inequalities of this and more general types and exploit them in limiting and guiding the enumerative search in solving zero–one problems. Their propagation procedure is used to fix variables or to finally give what we call a transitively closed graph. Hammer and Ngyen [8] used degree-two inequalities as a key in their APOSS procedure.

We will represent these systems of inequalities by a certain type of graph, called bigraphs, which are similar to the bidirected graphs for the general matching problem [4].

A case of special interest is the independent node problem. In that problem, a zero–one variable x_i is associated with node i, for each node i, and the inequality $x_i + x_j \leq 1$ is required if the edge $[i, j]$ is present. The nodes i having $x_i = 1$ and satisfying these inequalities form an independent set of nodes, i.e., a set of nodes such that no edge connects any two nodes in the set. Any set packing problem [11] can formally be converted to an independent node problem on the 'intersection graph', i.e., the graph with a node i for the ith subset and an edge $[i, j]$ if the ith and jth subset have a nonempty intersection.

The node covering problem involves constraints $x_i + x_j \geq 1$ for every edge $[i, j]$. The nodes with $x_i = 1$ form a node cover, i.e., every edge meets at least one such node. The complement of a node cover is an independent set, and clearly a node cover containing the minimum number of nodes among all node covers for a given graph is the complement of an independent set containing the maximum number of nodes among all independent sets of the given graph.

Although a general set packing problem can be converted to a node packing problem via the intersection graph, the same is not true of set covering. In the case of set packing, there are many families of sets which will give the same intersection graph. The intersection graph, and its associated cliques to be discussed in Section 3, gives a canonical representation of any set packing problem.

2. Degree-two inequalities and bidirected graphs

A bidirected graph $G = (N, E)$ is a set N of *nodes* and a set E of *edges* where each edge $e \in E$ has two *ends*. Each *end* of an edge *meets* a node $i \in N$ and has an associated sign. The signs for the two ends of an edge need not agree. Thus, an edge e is denoted as the unordered pair $[i, j]$ where i and j are the two nodes met by the ends of edge e. In addition, the signs of the ends of the edge lead to three types of edges:

$(+,+)$ edges $e=[i,j]$ with two plus ends,

$(-,-)$ edges $e=[i,j]$ with two minus ends,

$(+,-)$ edges $e=[i,j]$ with a plus end at i and a minus end at j.

A $(-,+)$ edge of a bidirected graph G is defined accordingly. Note that edges of the form $[i,i]$ may present in the graph. We call such edges *loops*.

This definition is the same as in [4]. Here, however, we have a different use of such graphs. A system of inequalities each of which has two zero–one variables can be represented by a bidirected graph having a node i for each variable x_i and an edge for each inequality:

$$\begin{aligned} x_i + x_j &\geq 1 && \text{gives a } (+,+) \text{ edge } [i,j], \\ -x_i - x_j &\geq -1 && \text{gives a } (-,-) \text{ edge } [i,j], \\ x_i - x_j &\geq 0 && \text{gives a } (+,-) \text{ edge } [i,j]. \end{aligned}$$

Thus, for any inequality system with inequalities of these types there is a unique bidirected graph, and any bidirected graph represents such an inequality system.

A loop $[i,i]$ of a $(+,+)$ type corresponds to the inequality $x_i + x_i \geq 1$. In 0–1 variables, this inequality implies $x_i = 1$. Similarly, a $(-,-)$ loop gives $-x_i - x_i \geq -1$, or $x_i = 0$. A $(+,-)$ loop does not give an inequality since it would give $x_i - x_i \geq 0$.

Two edges e and e' may meet the same pair of nodes $[i,j]$ if they are of different type, e.g., if e is a $(+,+)$ edge and e' is a $(+,-)$ edge. That is, an inequality system may well include the two distinct inequalities:

$$x_i + x_j \geq 1 \quad \text{and} \quad x_i - x_j \geq 0\,.$$

Note that the last two inequalities imply a $(+,+)$ loop at node i.

Given a bidirected graph G, we form the *transitive closure* G^* of G, inductively, by adjoining edges to G as follows: if node j has two distinct incident edges e and e' with a plus end of e meeting j and a minus end of e' meeting j, then the new edge, if not already in G, to be adjoined to G has as its two ends the other end of e and the other end of e'. To be more explicit, suppose the edge e has the other end meeting node i and edge e' has the other end meeting node k. Then, the new edge e^* has one end meeting node i, with the same sign as the end of e meeting i, and the other end of e^* meets node k with the same sign as the end of e' meeting k (cf. Fig. 1). It is allowed that i be equal to k or that i be equal to j or that j be equal to k, but edges e and e' are required to be distinct.

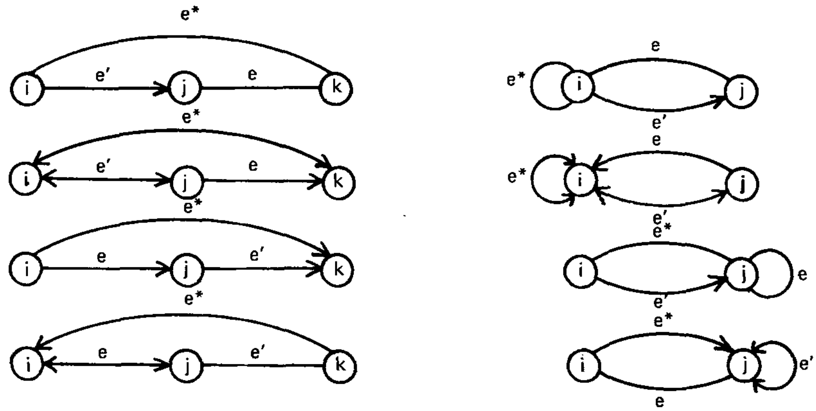


Fig. 1.

We need not adjoin an edge already present *with* the same sign pattern because that would correspond to writing the same inequality twice. Similarly, a $(+,-)$ loop can be dropped because it corresponds to the trivial inequality $x_i - x_i \geqslant 0$.

A *transitively closed* graph is a bidirected graph G such that G is equal to its transitive closure G^*. That is, all of the edges e^* which might be required to be adjoined to G are already present in G.

A *simple* bidirected graph is a bidirected graph with no loops and no two edges meeting the same pair of nodes.

To obtain a simple, transitively closed graph from a transitively closed graph we carry out the following reductions. (There is no need to repeat the process of forming the transitive closure on the reduced graph.)

First, if a node i has a $(+,+)$ loop, then the inequality $x_i + x_i \geqslant 1$ must be satisfied, so any 0–1 solution must have $x_i = 1$. Likewise, if $(-,-)$ loop is at node i, then $-x_i - x_i \geqslant -1$, so $x_i = 0$ is satisfied by every 0–1 solution. In this case, the node i and all incident edges can be deleted from the graph. Clearly, if a node i has both a $(+,+)$ loop and a $(-,-)$ loop, then there is no 0–1 solution to the problem. In the next section we prove that there is a 0–1 solution whenever the transitive closure has no node with both a $(+,+)$ loop and a $(-,-)$ loop.

We also show that if the transitive closure has no loops at all, then for any node i there is a 0–1 solution with $x_i = 0$ and another 0–1 solution with $x_i = 1$.

Second, if there is a $(+,-)$ edge $[i,j]$ and a $(-,+)$ edge $[i,j]$, then $x_i \geqslant x_j$ and $x_j \geqslant x_i$ so $x_i = x_j$ and we eliminate one variable. In the graph, we 'shrink' the two nodes $[i,j]$ together as in Fig. 2(a). Likewise, if there is a $(+,+)$ edge $[i,j]$ and a $(-,-)$ edge $[i,j]$, then $x_i + x_j \geqslant 1$ and $-x_i - x_j \geqslant -1$ hold and thus $x_i = 1 - x_j$. We can eliminate x_i using the substitution $x_i = 1 - x_j$. Fig. 2(b) illustrates this case.

Third, if there is a $(+,-)$ edge $[i,j]$ and a $(+,+)$ edge $[i,j]$, then we have a

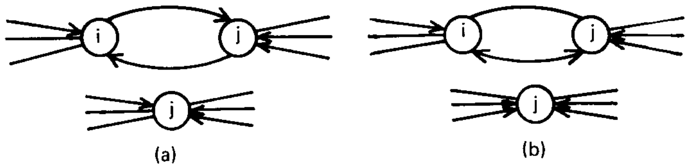


Fig. 2.

(+, +) loop for node i. Likewise, if there is a (+, −) edge $[i, j]$ and a (−, −) edge $[i, j]$, we have a (−, −) loop for node j. This shows that we can always work with simple, transitively closed graphs.

We are interested in these graphs because we want to represent inequality systems. Thus, we can consider that a duplicate (+, +) edge $[i, j]$ can be eliminated in the same way that an inequality $x_i + x_j \geq 1$ can be eliminated if it has already appeared in the inequality system. In the same way, changing a variable x_i to $x'_i = 1 - x_i$ does not essentially change the inequality system. This change is represented in the graph by changing all the signs of ends of edges at a given node. That is, changing x_i to $1 - x_i$ changes each plus end meeting node i to a minus end and each minus end meeting node i to a plus end without affecting the transitive closure of the bidirected graph.

Therefore, we can say that our graphs are not essentially changed by duplicating an edge, including the signs of its ends, or by changing all of the signs at a node.

The construction of the transitive closure of a given bidirected graph G can be done by 'scanning' each edge of the transitive closure G^* exactly once. We first state what the main step, scanning an edge, consists of.

Scan edge $e = [i, j]$: If e has a plus end at node i, look at every other edge e' with a minus end at i and adjoin e^*, as required in the definition of transitive closure, if it is not already present. If e has a minus end at node i, then look at every other edge e' with a plus end at i. Do the same for node j.

We can start with a list of edges in any order and start scanning each edge in turn. New edges e can be adjoined to the bottom of the list. Terminate when every edge, either originally present or adjoined, has been scanned.

To show that this simple scanning procedure produces the transitive closure requires only showing that, for the graph produced, every pair of edges e and e' meeting node i with oppositely signed ends eventually gets looked at. But scanning the edge e or e' further down on the list must result in looking at the pair because at that point the other edge is present.

Since the order of work in the scanning step can be kept down to $O(n)$ and the number of edges in G^* is at most $O(n^2)$, the total work required to get the transitive closure of a bidirected graph is at most $O(n^3)$, where $n = |N|$.

3. Bigraphs

For this section we concern ourselves with simple, transitively closed bidirected graphs. Call such graphs *bigraphs*. A 0–1 solution to the inequality system associated with a bigraph G is called a 0–1 *solution for* G.

Proposition 1. *The subgraph G_S induced by a subset S of the nodes of a bigraph G is a bigraph.*

Proof. Clearly G_S is bidirected and simple. G_S is transitively closed because every edge of the transitive closure of G_S is an edge of the transitive closure of G. Since G is a bigraph the proposition follows.

The following two constructions are used frequently in both this section and the next.

Construction 0 (Assigning a variable the value 0) (cf. Fig. 3). For any bigraph G let T be the subset of nodes of G consisting of a node i and all of its neighboring nodes j met by an edge $[i, j]$ having a plus end at node i and let $R = N - T$. For $k \in T$ we construct a partial solution x^T as follows:

$$x_k = \begin{cases} 0 & \text{if } k = i \text{ or } [i, k] \text{ is a } (+, -) \text{ edge,} \\ 1 & \text{if } [i, k] \text{ is a } (+, +) \text{ edge.} \end{cases}$$

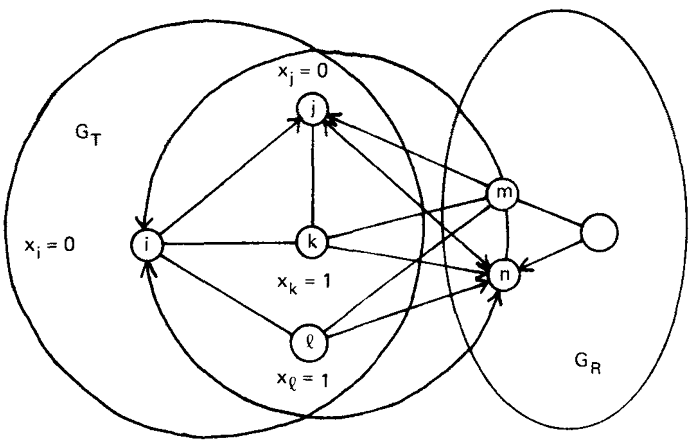


Fig. 3. Construction 0.

Construction 1 (Assigning a variable the value 1) (cf. Fig. 4). For any bigraph G let U be the subset of nodes of G consisting of a node i and all of its neighboring nodes j met by an edge $[i, j]$ having a minus end at node i and let $S = N - U$. For $k \in U$ we construct a partial solution x^U as follows:

$$x_k = \begin{cases} 1 & \text{if } k = i \text{ or } [i, k] \text{ is a } (-, +) \text{ edge.} \\ 0 & \text{or } [i, k] \text{ is a } (-, -) \text{ edge.} \end{cases}$$

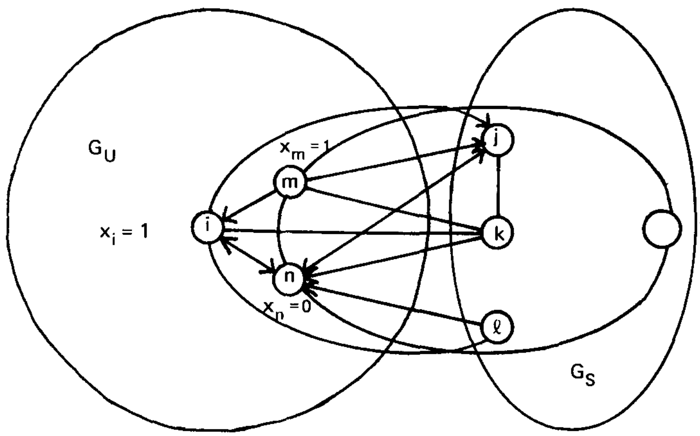


Fig. 4. Construction 1.

Proposition 2. *Let G be a bigraph and let $i \in N$, T, R, S and U be defined as in Construction* 0 *and Construction* 1, *respectively.*

(i) *If x^R is a* 0–1 *solution for the subgraph G_R, then $x = (x^T, x^R)$ is a* 0–1 *solution for G.*

(ii) *If x^S is a* 0–1 *solution for the subgraph G_S, then $x = (x^U, x^S)$ is a* 0–1 *solution for G.*

Proof. We prove part (i) of the proposition; the proof of part (ii) is similar. By construction x^T is a 0–1 solution for G_T. Hence it suffices to show that the inequalities corresponding to edges $e = [j, k]$ with $j \in T$ and $k \in R$ are satisfied by x^T and, more precisely, that they are rendered superfluous by the choice of x^T. We consider the four cases that are possible and show that only two of them can occur in the bigraph G:

Case 1. $x_j = 0$ and edge e has a plus end at j;
Case 2. $x_j = 1$ and edge e has a plus end at j;
Case 3. $x_j = 0$ and edge e has a minus end at j;
Case 4. $x_j = 1$ and edge e has a minus end at j.

In Case 2 the corresponding inequality is either $x_j - x_k \geqslant 0$ or $x_j + x_k \geqslant 1$. With x_j fixed to the value 1, the inequality is satisfied and reduces to $x_k \leqslant 1$ or $x_k \geqslant 0$. Thus in Case 2 the assertion follows. In Case 3 the assertion follows by an analogous argument.

Suppose now Case 1 occurs. Node j cannot be equal to the node i chosen in Construction 0 because otherwise $k \in T$ holds. Since $j \in T$, $j \neq i$ and $x_j = 0$ hold, the edge $[i, j]$ is a $(+, -)$ edge. Since the edge $e = [j, k]$ has a plus end at j and since G is transitively closed, it follows that the edge $e^* = [i, k]$ is in G and that it has a plus end at node i. Hence, $k \in T$ follows, contradicting $k \in R = N - T$. Thus, Case 1 cannot occur. By analogous reasoning, Case 4 cannot occur. Hence, we are either in Case 2 or 3 and taking any 0–1 solution x^R for G_R yields a 0–1 solution for G when combined with x^T. Proposition 2 follows.

We remark that this proposition is true only because G is transitively closed, which means intuitively, that the implications of setting $x_i = 0$ (or 1) have already been propagated through the graph, see [7].

Proposition 2 does not assure that there is a 0–1 solution for G because there may be no 0–1 solution for G_R. However, the existence of a 0–1 solution for a bigraph G is easily proven from Proposition 2 by induction. In fact, more is true.

Proposition 3. *Let G be a bigraph and $k \in N$ be any node of G. Then there exist 0–1 solutions x^0 and x' for G such that $x_j^0 = x_j'$ for all $j \neq k$, $x_k^0 = 0$ and $x_k' = 1$ hold.*

Proof. Let W be the set of nodes of G which are not connected to node k by an edge of G. We apply, e.g., Construction 0 to any node $i \in W$ and continue to do so until we are left with a subgraph G' with node set N' such that $k \in N'$ and every node of G' is linked to the node k by some edge of G'. If $i \in N'$, $i \neq k$, is linked to node k by an edge having a plus end at node i we set $x_i = 1$ and use Construction 1. It follows that node k is in the set S that results and thus remains unfixed. If $i \in N'$, $i \neq k$, is linked to node k by an edge having a minus end at node i we set $x_i = 0$ and use Construction 0. It follows that node k is in the set R that results and thus remains unfixed. In either case we can apply Construction 0 or 1 again and repeat until node k is an isolated node. Thus x_k can be assigned arbitrarily the value 0 or 1 and Proposition 3 follows.

Proposition 4. *Let G be a bigraph and let $i \neq j$ be any two nodes of G. For any assignment of 0–1 values to x_i and x_j which is not excluded by an edge of G (if present) there exists a 0–1 solution for G.*

Proof. Consider first the case where there is no edge $[i, j]$ in G. Then, we prove the proposition by first applying Construction 0 or 1, as required, for node i and following with Construction 0 or 1, again as required, for node j, which will be in G_R or G_S, as the case may be. Since there is no edge $[i, j]$, Construction 0 (or 1) for node i will leave node j in G_R (or G_S) so that either construction can then be applied to node j.

To prove the proposition, in general, requires considering several cases. We consider the case of a $(+, +)$ edge $e = [i, j]$ and leave the other cases to the reader. The values excluded by a $(+, +)$ edge for $[i, j]$ are $x_i = 0$ and $x_j = 0$ since the corresponding inequality is $x_i + x_j \geqslant 1$.

Thus, we need to show that all three of $x_i = 0$ and $x_j = 1$, $x_i = 1$ and $x_j = 0$, $x_i = 1$ and $x_j = 1$ can occur in 0–1 solutions for G. If Construction 0 is applied to node i, then a 0–1 solution for G having $x_i = 0$ and $x_j = 1$ will be produced. Applying Construction 1 to node i leaves node j in G_S so that either $x_j = 0$ or $x_j = 1$ is possible in a 0–1 solution for G_S. Using Proposition 2, there are 0–1 solutions for G with $x_i = 1$ and x_j equal to either 0 or 1, completing the proof.

Proposition 4 shows that the transitive closure of G gives all of the possible implications on pairs of variables taking on 0–1 values. This result is even stronger than saying that there are no new inequalities on pairs of variables which can be derived as non-negative combinations of the existing inequalities.

Theorem 5. *The convex hull of* 0–1 *solutions for a bigraph* G *is a full-dimensional polytope in* $\mathbb{R}^n$.

The proof of Theorem 5 follows directly from Proposition 3.

4. Clique facets

A *clique* in a graph is a maximal, completely connected subgraph of the graph. Thus, the set S of nodes of the clique has the property that for any pair of nodes in S, there is an edge of G meeting that pair of nodes; and adjoining another node of G to S would cause that completely connected property to no longer hold.

Let G be a bigraph with node set N and let G_S be a completely connected subgraph of G with node set S. Define $S_+ \subseteq S$ ($S_- \subseteq S$, respectively) the subset of nodes of G_S which are met by an edge of G_S having a plus end (a minus end respectively). Clearly, $S = S_+ \cup S_-$ holds, but it is possible that $S_+ \cap S_- \neq \emptyset$. We call G_S (or simply, S) a *biclique* in G if:

(i) G_S is completely connected;

(ii) $S_+ \cap S_- = \emptyset$ holds;

(iii) G_S is a maximal subgraph of G with respect to the requirements (i) and (ii).

For the bigraph shown in Fig. 5, the nodes 1, 2, 3 form a biclique with $S_+ = \{1, 3\}$ and $S_- = \{2\}$. Although nodes 1, 2, 3, 4 form a clique, they do not form a biclique because node 4 is met by both plus ends and minus ends of edges connecting node 4 to nodes 1, 2, 3. Similarly, nodes 1, 2, 3, 5 form a clique but not a biclique because now node 2 is met by both plus ends and minus ends of edges connecting it to nodes 1, 3, 5. Nodes 1, 2, 3, 6 do not form a biclique because they are not completely connected. Thus, nodes 1, 2, 3 are maximal with respect to the required property and do form a biclique.

A biclique of G with node set $S = S_+ \cup S_-$ is called a *strong biclique* of G if there does not exist a node $k \notin S$ with edges $[k, i]$ for all $i \in S$ in G such that the edge $[k, i]$ has a plus end meeting i if $i \in S_+$ and a minus end meeting i if $i \in S_-$. In Fig. 5, nodes 1, 2, 3 do not form a strong biclique because of node 4. The strong bicliques in the graph shown in Fig. 5 are $\{1, 3, 4\}$, $\{1, 3, 6\}$, $\{2, 4\}$, and $\{2, 5\}$.

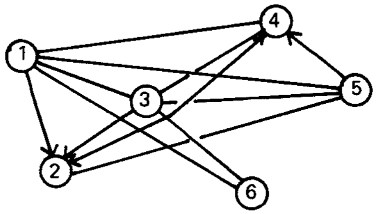


Fig. 5.

For a bigraph G and a biclique of G with nodes $S = S_+ \cup S_-$, define the corresponding *clique inequality* to be the inequality

$$\sum_{j \in S_+} x_j - \sum_{j \in S_-} x_j \geq |S_+| - 1 .$$

Proposition 6. *Every* 0–1 *solution for a bigraph G satisfies the clique inequalities of G.*

Proof. The clique inequality for $S = S_+ \cup S_-$ can be written as

$$\sum_{j \in S_+} (1 - x_j) + \sum_{j \in S_-} x_j \leq 1 .$$

In order to violate it, there must be one of the following possibilities:

(i) $x_i = 1$ and $x_j = 1$ for $i \in S_-$ and $j \in S_-$;
(ii) $x_i = 1$ and $x_j = 0$ for $i \in S_-$ and $j \in S_+$;
(iii) $x_i = 0$ and $x_j = 0$ for $i \in S_+$ and $j \in S_+$.

Each of these three possibilities is explicitly excluded by an edge of G, as illustrated in Fig. 6.

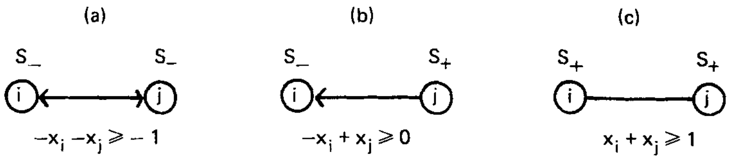


Fig. 6.

Proposition 7. *For a bigraph G and a biclique with node set $S = S_+ \cup S_-$ of G, if T is a subset of S, then the inequality*

$$\sum_{j \in T_+} x_j - \sum_{j \in T_-} x_j \geqslant |T_+| - 1\,,$$

where $T_+ = T \cap S_+$ and $T_- = T \cap S_-$, is satisfied by every 0–1 solution for G. However, this inequality for T is implied by the clique inequality corresponding to S and the linear inequalities $0 \leqslant x_j \leqslant 1$, for all $j \in N$.

Proof. The fact that every 0–1 solution satisfies this inequality corresponding to T has the same proof as for Proposition 6.

The second assertion is shown by adding the inequalities

$$\sum_{j \in S_+} x_j - \sum_{j \in S_-} x_j \geqslant |S_+| - 1\,,$$
$$-x_j \geqslant -1, \quad j \in S_+ - T_+\,,$$
$$x_j \geqslant 0, \quad j \in S_- - T_-\,,$$

to derive the inequality corresponding to T.

Proposition 8. *If a biclique S of a bigraph G is not a strong biclique, then there exist two bicliques $C_1 \neq S$ and $C_2 \neq S$ such that the clique constraint corresponding to S is implied by the two clique inequalities corresponding to C_1 and C_2. Thus, the clique constraint for S is implied by the system of strong biclique inequalities.*

Proof. Suppose the biclique S with nodes $S = S_+ \cup S_-$ is not strong, i.e., there exists $k \notin S$ such that the edge of $[k, i]$ is in G for every $i \in S$ and has a plus end (a minus end, respectively) at the node i when $i \in S_+$ ($i \in S_-$). By the maximality of S, the node k must have some plus ends and some minus ends meeting k for edges $[k, i]$, $i \in S$. Define

$$T_+ = \{i \in S \mid \text{edge } [k, i] \text{ has a plus end meeting } k\},$$

$$T_- = \{i \in S \mid \text{edge } [k, i] \text{ has a minus end meeting } k\}.$$

It is clear that $C_1 = \{k\} \cup T_+$ and $C_2 = \{k\} \cup T_-$ are bicliques with

$$C_1 = [\{k\} \cup (S_+ \cap T_+)] \cup [S_- \cap T_+] = C_1^+ \cup C_1^-,$$

$$C_2 = [S_+ \cap T_-] \cup [\{k\} \cup (S_- \cap T_-)] = C_2^+ \cup C_2^-,$$

being partitions of C_1 and C_2 so that their clique inequalities are

$$\sum_{j \in C_1^+} x_j - \sum_{j \in C_1^-} x_j \geq |C_1^+| - 1 \quad \text{and} \quad \sum_{j \in C_2^+} x_j - \sum_{j \in C_2^-} x_j \geq |C_2^+| - 1.$$

Adding these two inequalities gives

$$\sum_{j \in S_+} x_j - \sum_{j \in S_-} x_j \geq 1 + |S_+| - 2 = |S_+| - 1,$$

because x_k appears with $a + 1$ and $a - 1$ in the two inequalities. Hence, the first assertion is proven.

To prove that the system of strong biclique inequalities suffices, let P denote the polytope defined by all clique constraints and the constraints $0 \leq x_j \leq 1$, $j \in N$. By Theorem 5, P has dimension n because P contains the convex hull of 0–1 solutions. Consequently, every facet of P has dimension $n - 1$. Let $ax \geq a_0$ be an inequality corresponding to a biclique which is not strong. By the first part of the proof, there exist two clique inequalities $cx \geq c_0$ and $dx \geq d_0$ such that $a = c + d$ and $a_0 = c_0 + d_0$. Hence, $ax = a_0$ if and only if $cx = c_0$ and $dx = d_0$. Furthermore, the $2 \times n$ matrix with rows c and d has rank 2. It follows that $ax = a_0$ defines a face of P of dimension at most $n - 2$. The second assertion of the proposition is, thus, proven.

Propositions 6, 7 and 8 prove the following result.

Theorem 9. *For a bigraph G the set of 0–1 solutions to the inequality system*

$$0 \leqslant x_j \leqslant 1, \quad j \in N,$$
$$\sum_{j \in T_+} x_j - \sum_{j \in T_-} x_j \geqslant |T_+| - 1 \quad \textit{for all } T \subseteq C \textit{ and for all } C, \; C \textit{ a biclique in } G \tag{4.1}$$

are exactly the 0–1 *solutions for* G. *Furthermore, the inequality system*

$$0 \leqslant x_j \leqslant 1, \quad j \in N,$$
$$\sum_{j \in C_+} x_j - \sum_{j \in C_-} x_j \geqslant |C_+| - 1 \quad \textit{for all strong bicliques } C \textit{ in } G \tag{4.2}$$

has the same (*real*) *solutions* $x \in \mathbb{R}^n$ *as the* (*in general, considerably larger*) *system* (4.1).

Proof. By Propositions 6 and 7 we have that every 0–1 solution for G satisfies (4.1). On the other hand, since every edge of G is contained in some biclique of G it follows by the second part of Proposition 7 that every 0–1 solution to (4.1) is a 0–1 solution for G. Again by Proposition 7 it follows that in the inequality system (4.1) we can restrict ourselves to considering bicliques only without changing the real solution space. By Proposition 8, if a biclique C is not a strong biclique, then the resulting inequality is implied by the clique inequalities associated with the strong bicliques. Thus, Theorem 9 follows.

When the bigraph G has only $(-,-)$ edges the original inequalities are all of the type $-x_i - x_j \geqslant -1$, the strong bicliques are cliques in the usual meaning (and vice versa), and the form of the clique inequality system is of the form related to graphs. With respect to graphs it is known [5, 10] that all clique inequalities define facets of the associated 0–1 polytope.

For general bigraphs, we can ask the question whether any strong biclique inequality can be omitted without changing the solution set in $\mathbb{R}^n$. The next theorem answers the question in the negative.

Theorem 10. *The strong biclique inequalities are facets of the convex hull of* 0–1 *solutions for the bigraph* G.

Proof. The proof follows the lines of the proof for undirected graphs [10] but is more difficult.

We must show, by Theorem 5, that there exist n affinely independent 0–1 solution vectors satisfying the strong biclique inequality with equality. Let $C = C_+ \cup C_-$ be the node set of a strong biclique and let $c_1 = |C_+|$, $c_2 = |C_-|$ and $c = |C|$. Assume without loss of generality that $C = \{1, \ldots, c\}$. Construct c_1 affinely independent 0–1 solutions for G by setting $x_i = 0$ for exactly one $i \in C_+$

at a time using Construction 0. By Proposition 3 such a 0–1 solution exists. Construct c_2 affinely independent 0–1 solutions for G by setting $x_i = 1$ for exactly one $i \in C_-$ at a time using Construction 1 and invoking Proposition 3. The resulting c 0–1 solutions for G satisfy the clique inequality with equality and are affinely independent among themselves as follows from the construction of the $c \times c$ matrix on the first c columns corresponding to the nodes in C.

To complete the proof we construct $2(n-c)$ 0–1 solutions for G satisfying the inequality with equality as follows: Since C is a strong biclique for each node $k \not\in C$ there exist at least one node $i \in C$ such that either $i \in C_+$ and there is no edge $[i, k]$ with a plus end at i, or $i \in C_-$ and there is no edge $[i, k]$ with a minus end at node i. In the former case, set $x_i = 0$ and use Construction 0; in the latter case, set $x_i = 1$ and use Construction 1. In either case, the construction fixes all x_j with $j \in C$ so that the clique inequality is satisfied with equality and so that the node k satisfies $k \in R$ in Construction 0 or $k \in S$ in Construction 1. By Proposition 1, the resulting subgraph G_R (G_S, respectively) is a bigraph and we can apply Proposition 3 to G_R (G_S, respectively). Combining the statements it follows that for each $k \not\in C$ there exist two solutions x^0 and x^1 each satisfying the clique inequality with equality and such that $x_j^0 = x_j^1$ for all $j \neq k$, $x_k^0 = 0$ and $x_k^1 = 1$ hold. We list the $2(n-c)$ 0–1 solutions for G thus obtained pairwise and observe that by elementary row operation we obtain a $(n-c) \times (n-c)$ identity matrix in columns $c+1, \ldots, n$. Consequently, by construction, we have n affinely independent 0–1 solutions to G satisfying the clique inequality with equality and thus Theorem 10 follows.

We note that—like in the case of set-packing polyhedra, see [10, 11, 13]—it is natural to look for other facet-defining structures in a bigraph G—generalizing the notion of odd cycles, webs, etc. Another avenue for research is a generalization of the 'lifting procedure' for facets to the case of bigraphs. These possibilities are at present left for future research.

5. Biperfect graphs

In this section, we define the notion of biperfect graphs, give some examples and simple results, and make several conjectures. The principal conjecture is that the class of biperfect graphs is really the same as the class of perfect graphs.

For a given bigraph G define its *clique matrix* M to be the matrix whose rows consist of the coefficient of the strong biclique inequalities. Define the *clique polytope* of G to be the set of $x \in \mathbb{R}^n$ satisfying $0 \leqslant x \leqslant 1$ and satisfying the clique inequalities for strong bicliques of G. We are interested in the question of when every vertex of the clique polytope is a 0–1 vector.

Another way to view the clique inequalities is to start with an inequality system of the form

$$0 \leq x_j \leq 1, \quad j = 1, \ldots, n,$$

$$\sum_{j=1}^{n} a_{ij} x_j \geq p_i - 1, \quad i = 1, \ldots, m,$$

where every a_{ij} is 0, 1 or -1 and p_i is equal to the number of $+1$'s among $a_{i1}, \ldots, a_{in}$. When all a_{ij} are 0 and -1, we have a set packing problem, and we wish to generalize the clique approach and perfect graph framework [5] for that problem. Given an inequality system with 0, 1, -1 coefficients we can form a bidirected graph G by making a biclique among the nodes $N = \{1, 2, \ldots, n\}$ for each row of $A = (a_{ij})$. That is, put in G an edge $[j, k]$ of type (a_{ij}, a_{ik}) whenever there is a row i with $a_{ij} \neq 0$ and $a_{ik} \neq 0$. Then the 0–1 solutions for this bidirected graph G are the same as the 0–1 solutions for the original system. The particular right-hand side, $p_i - 1$, required in the original system is critical here. The original inequality system cannot have a polytope with only 0–1 vertices unless every strong biclique inequality is among the rows of A, that is, unless the clique matrix M of G is included as a submatrix of A. In general, that inclusion is not sufficient to assure 0–1 vertices, but the clique matrix M does give a smaller polytope of solutions, or a tighter linear programming relaxation, than the original system of linear inequalities.

We could thus define biperfect graphs in terms of *biperfect* matrices which yield polytopes having 0–1 extreme points only but we instead follow the classical approach for perfect graphs [5].

First, let us discuss what it means to bidirect an undirected graph G^0. An undirected graph has no sign, + or −, assigned to the ends of edges. To *bidirect* G^0 means to assign a plus or minus to each end of each edge, giving edges which may be of type $(+, +)$, $(+, -)$, $(-, +)$ or $(-, -)$. To *bidirect* G^0 *to form a bigraph* G means to bidirect G^0 to give a bigraph G, that is, to give a bidirected graph which is simple (which is sure if G^0 is simple) and which is equal to its transitive closure. The latter condition is rather stringent since, for example, we can only bidirect a graph G^0 to a bigraph G having only $(+, -)$ or $(-, +)$ edges if G^0 is a comparability graph [6]. Any undirected graph G^0 can, however, always be given some bidirection just by making every end a plus end or by making every end a minus end. Then, the resulting bidirected graph G is a bigraph provided only that G^0 was simple. We allow other bidirections, one could say, in between these two extreme cases.

Given a bigraph G, we consider optimization problems of the form

$$x_j = 0 \text{ or } 1, \quad j = 1, \ldots, n,$$

$$\sum_{j \in S_+} x_j - \sum_{j \in S_-} x_j \geq |S_+| - 1, \quad \text{for all strong bicliques } S = S_+ \cup S_-,$$

$$\sum_{j \in Z_+} x_j - \sum_{j \in Z_-} x_j = z \quad \text{(minimize)},$$

where Z_+ and Z_- are disjoint subsets of N. In other words, we consider optimizing 0–1 solutions for G with objective functions $z = \Sigma\, c_j x_j$ having $c_j = 0$, $+1$, or -1. Then,

$$Z_+ = \{j \mid c_j = +1\}, \qquad Z_- = \{j \mid c_j = -1\}, \qquad Z_0 = \{j \mid c_j = 0\},$$

and Z_0, Z_+, Z_- form a partition of the nodes.

The linear programming relaxation replaces $x_j = 0$ or 1 by $0 \leq x_j \leq 1$. We form the dual linear program and consider the optimization problem of finding integer answers to it. This dual problem has a variable π_S for each strong biclique and a variable δ_j for each node. The constraints are

$$\pi_S \geq 0 \text{ and integer} \quad \text{for all strong bicliques } S,$$

$$\delta_j \geq 0, \quad j = 1, \ldots, n,$$

$$\sum_{S:\, j \in S_+} \pi_S - \sum_{S:\, j \in S_-} \pi_S - \delta_j \leq \begin{cases} 0, & j \in Z_0, \\ 1, & j \in Z_+, \\ -1, & j \in Z_-, \end{cases}$$

$$\sum_S (|S_+| - 1)\pi_S - \sum_{j=1}^{n} \delta_j = v \quad \text{(maximize)}.$$

The summations over S are sums over all strong bicliques $S = S_+ \cup S_-$. In general, the objective v of the dual problem satisfies $v \leq z$ for z the objective of the original problem for G. We say a bigraph G has the (strong) *max–min property* if these two objectives are equal, vide Fulkerson [5].

Define an undirected graph G^0 to be *biperfect* if every bigraph G formed by bidirecting G^0 satisfies the max–min property.

A biperfect graph is obviously perfect because one way of directing an undirected graph is to make all ends be minus ends, and then the max–min property becomes the famous relation (maximum number of independent nodes) = (minimum number of cliques needed to cover all nodes) [2].

We conjecture that a graph is biperfect if and only if it is perfect. Some reasons for thinking that the conjecture may be true will be given.

Proposition 11. *Bipartite graphs are biperfect.*

Proof. A bipartite graph has no triangles so it has no cliques other than the edges themselves. For a bipartite graph G^0, the dual linear program for some bidirection of G^0 is

$$\pi_e \geqslant 0 \text{ and integer,} \quad \text{for all edges } e,$$

$$\delta_i \geqslant 0, \quad \text{for all nodes } i,$$

$$\sum_{\substack{e \text{ meets } i \\ \text{with a + end}}} \pi_e - \sum_{\substack{e \text{ meets } i \\ \text{with a} - \text{end}}} \pi_e - \delta_i \leqslant \begin{cases} 0, & i \in Z_0, \\ 1, & i \in Z_+, \\ -1, & i \in Z_-, \end{cases}$$

$$\sum_{\substack{e \text{ has two} \\ + \text{ ends}}} \pi_e - \sum_{\substack{e \text{ has two} \\ - \text{ ends}}} \pi_e - \sum_{i=1}^{n} \delta_i = v \quad \text{(maximize)}.$$

A linear program of this type has integer answers to both its primal and dual whenever there does not exist any odd circuit [4], where here an odd circuit means a circuit with an odd number of edges having either two plus ends or two minus ends. We can obviously bidirect a bipartite graph with a circuit to have an odd circuit just by making all ends plus ends except one (cf. Fig. 7).

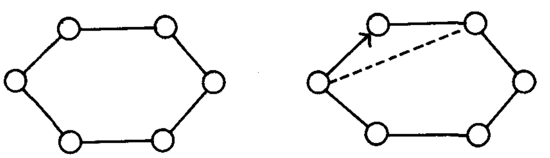


Fig. 7.

However, doing so causes a triangle to be formed in the transitive closure so that the resulting bidirected graph could not be transitively closed since we started with a bipartite graph, which has no triangles.

In general, any bidirection of a bipartite graph in order to be transitively closed will have to make both ends meeting a node in any circuit either both plus or both minus ends. Then, the resulting circuit will not be an odd circuit as can be seen by a simple parity argument. Therefore, any bipartite graph is biperfect.

This proof suggest the next proposition. First, define an odd-dihole to be a hole (i.e., a circuit with no chords) having an odd number of edges with either both plus ends or both minus ends. An odd hole in an undirected graph is a hole with an odd number of edges. Any bidirected graph with an odd-dihole does not satisfy the max–min property.

Proposition 12. *An undirected graph G^0 with an odd hole has an odd-dihole in every bigraph G formed by bidirecting G^0. Conversely, if a bigraph G has an odd-dihole, then the undirected graph G^0 formed by just dropping the + or − sign on each end of edges has an odd hole.*

The proof is essentially that already given in the proof of Proposition 11. What Proposition 12 says is that if a graph is not biperfect because some bigraph formed from it had an odd-dihole, then the original graph was not perfect because it had an odd hole. However, it says something in the other direction as well, namely, if a graph is not perfect because it has an odd hole, then every bigraph formed from it will not have the max–min property because of an odd-dihole.

We conjecture a stronger version of the previous conjecture: if some bigraph G formed from G^0 has the max–min property, then G^0 is biperfect. That is, all bigraphs formed from G^0 have the max–min property or none does.

An interesting example is given by comparability graphs. A comparability graph is an undirected graph such that directions can be assigned to each edge so that the resulting directed graph is acyclic and transitively closed. The ordering then given by $i > j$ if an edge has a plus end meeting i and a minus end meeting j is a partial order. Such graphs (the original undirected versions) are known to be perfect. The strong max–min property for such graphs says that the maximal number of pair-wise incomparable elements in a partial order is equal to the minimum number of chains (linearly ordered subsets) covering all elements. Although a comparability graph may have several partial orders which can be formed by directing the edges, the incomparable elements and the chains are the same in all such partial orders.

We do not know if comparability graphs are biperfect. We conjecture that they are. However, one way of bidirecting the edges (other than making all plus ends) that works is to direct the edges so as to give a partial order. The max–min property holds because the resulting matrix is totally unimodular, being a network flow matrix. The max–min property has an interesting statement here. First, define a node with a $+1$ cost coefficient to be a *source* and a node with a -1 cost coefficient to be a *sink*. Define an *upper dominated set* to be a set S of nodes such that if $j \in S$ and $i > j$ then $i \in S$. The primal problem is to find an upper dominated set with the largest surplus of sinks minus sources. The dual problem amounts to finding pairings of sources to sinks by directed edges so that each source is paired to only one sink and the fewest number of sinks is left unpaired. The max–min property is easy to show from linear programming duality and the total unimodularity of the coefficient matrix.

If perfect graphs are biperfect it suffices to show the max–min property for the directed version of the problem. The reason that this could not be used for graphs other than comparability graphs is that only comparability graphs give a directed version which is simple and transitively closed.

References

[1] E. Balas and R. Jerowslow, Canonical cuts in the hypercube, SIAM J. Appl. Math. 23 (1972) 661–669.
[2] C. Berge, Graphs and Hypergraphs (North-Holland, Amsterdam, 1973).
[3] V. Chvátal, On certain polytopes associated with graphs, J. Comb. Theory Ser. B 18 (1975) 138–154.
[4] J. Edmonds and E.L. Johnson, Matching: a well-solved class of integer programs, in: R. Guy, ed., Combinatorial Structures and their Applications (Gordon and Breach, New York, 1970).
[5] D.R. Fulkerson, Blocking and anti-blocking pairs of polyhedra, Math. Programming 1 (1971) 168–194.
[6] P.C. Gilmore and A.J. Hoffman, A characterization of comparability and of interval graphs, Canad. J. Math. 16 (1964) 539–548.
[7] M. Guignard and K. Spielberg, Propagation, penalty improvement and use of logical inequalities in interactive 0–1 programming, Int. Symp. Math. Prog., Heidelberg University, 1976.
[8] P.L. Hammer and S. Nguyen, APOSS—a partial order in the solution space of bivalent programs, Publ. Centre Rech. Math. 163, Univ. Montréal, 1972.
[9] L. Lovász, Normal hypergraphs and the perfect graph conjecture, Discrete Math. 2 (1972) 253–267.
[10] M.W. Padberg, On the facial structure of set packing polyhedra, Math. Programming 5 (1973) 199–215.
[11] M.W. Padberg, Covering, packing and knapsack problems, Ann. Discrete Math. 4 (1979) 265–287.
[12] K. Spielberg, Minimal preferred variable reduction for zero–one programming, Rept. 320-3013, IBM Phil. Sc. C. (1972).
[13] L.E. Trotter, Jr., A class of facet producing graphs for vertex packing polyhedra, Discrete Math. 12 (1975) 373–391.

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 189–200

FLOW NETWORK FORMULATIONS OF POLYMATROID OPTIMIZATION PROBLEMS*

E.L. LAWLER

Computer Science Division, University of California, College of Engineering, Berkeley, CA 94720, USA

C.U. MARTEL

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

In the 'polymatroidal' network flow model, capacity constraints are imposed by polymatroid rank functions on the sets of arcs directed into and out of each node. It is shown that a variety of matroid optimization problems can be easily formulated and solved in terms of this model. Among these problems are (poly)matroid intersection, matroid partitioning, problems of gammoids and linking systems, and problems formulated by Iri, Krogdahl, and Fujishige. It is shown that simple proofs of known min–max theorems for these problems are easily obtained as corollaries of the max-flow min-cut theorem for polymatroidal network flows, previously proved by the authors.

1. Introduction

In the 'classical' network flow model, flows are constrained by the capacities of individual arcs. In the 'polymatroidal' network flow model, flows are constrained by the capacities of *sets* of arcs, where these capacities are imposed by polymatroid rank functions on the sets of arcs directed into and out of each node. Yet, as the authors have shown in another paper [10] the essential features of the classical model are retained; the augmenting path theorem, the integral flow theorem, and the max-flow min-cut theorem all yield to straightforward generalization. Moreover, a maximal integral flow can be computed efficiently, provided that there is a 'feasibility' oracle available for each capacity function.

We believe that the polymatroidal network flow model provides a satisfying generalization and unification of both network flow theory and much of the theory of (poly)matroid optimization. In this paper we demonstrate the usefulness of the model by providing network flow formulations of (poly)matroid

* This research was supported by National Science Foundation Grant MCS 78-20054.

intersection, matroid partitioning, problems of gammoids and linking systems, and problems formulated by Iri, Krogdahl and Fujishige. In fact, virtually every known problem in (poly)matroid optimization known to the authors can be easily formulated in terms of their network flow model, with the conspicuous exception of the polymatroid matching problem, solved by Lovász for the case of linearly represented polymatroids [11]. (Polymatroid matching cannot be formulated as a polymatroidal network flow problem, just as (nonbipartite) graphic matching cannot be formulated as an ordinary network flow problem.)

We assert that, when the maximal flow algorithm described in [10] is applied to each of the networks described in this paper, it either specializes to a known algorithm or is competitive with special algorithms which have been developed for the problems in question. We shall not elaborate on this point, but instead emphasize the ease with which min–max theorems can be proved, as corollaries of the general max-flow min-cut theorem proved in [10].

Remark. The polymatroidal network flow model described in this paper was formulated independently by Hassin [7], who considered a more general model in which costs are associated with flows in individual arcs and in which supermodular set functions impose lower bounds on flows through subsets of arcs. In [7] Hassin proved a (circulation) version of the max-flow min-cut theorem and also a more general version of the integrality theorem stated here. (These theorems require that the submodular upper bounds on flow and the supermodular lower bounds satisfy certain conditions.) He also developed algorithms for finding optimal flows, but gave no complexity estimates.

2. The polymatroidal network flow model

A *polymatroid* (E, ρ) is defined by a finite set of *elements* E and a *rank function* $\rho : 2^E \to \mathbb{R}^+ \cup \{\infty\}$ satisfying the following properties:

$$\rho(\emptyset) = 0\,, \tag{2.1}$$

$$\rho(X) \leqslant \rho(Y) \quad (X \subseteq Y \subseteq E)\,, \tag{2.2}$$

$$\rho(X \cup Y) + \rho(X \cap Y) \leqslant \rho(X) + \rho(Y) \quad (X \subseteq E,\ Y \subseteq E)\,. \tag{2.3}$$

Inequality (2.2) states that the rank function is monotone and (2.3) asserts that it is submodular. If also ρ is integer-valued and $\rho(\{e\}) = 0$ or 1 for all $e \in E$, then the polymatroid is a *matroid.* If (E, ρ) is matroid and $I \subseteq E$ is such that $|I| = \rho(I)$, then I is an *independent* set.

A *polymatroidal flow network* is a directed multigraph with a *source* s, a *sink* t

and two *capacity functions* α_j and β_j for each node j. Each function α_j (β_j) satisfies properties (2.1)–(2.3) with respect to the set of arcs A_j (B_j) directed out from (into) node j. Thus (A_j, α_j) and (B_j, β_j) are polymatroids. (Comment: We permit multiple arcs from one node to another. Hence A_j and B_j may be arbitrarily large finite sets.)

A *flow* in the network is an assignment of real numbers to the arcs of the network. Thus a flow is given by a function $f: E \to \mathbb{R}$, where E is the set of arcs. Such a function can be extended to subsets of arcs in a natural way, i.e.,

$$\begin{aligned} &f(\emptyset) = 0\,, \\ &f(X) = \sum_{e \in X} f(e) \quad (\emptyset \neq X \subseteq E)\,. \end{aligned} \tag{2.4}$$

(We continue to write $f(e)$ instead of the more cumbersome $f(\{e\})$.)

A flow is said to be *feasible* if

$$f(A_j) = f(B_j)\,, \quad j \neq s, t\,, \tag{2.5}$$

$$f(X) \leqslant \alpha_j(X) \quad \text{for all } j \text{ and } X \subseteq A_j\,, \tag{2.6a}$$

$$f(X) \leqslant \beta_j(X) \quad \text{for all } j \text{ and } X \subseteq B_j\,, \tag{2.6b}$$

$$f(e) \geqslant 0 \quad \text{for all arcs } e\,. \tag{2.7}$$

Eq. (2.5) imposes the customary flow conservation law at each node other than the source and sink. Inequalities (2.6a) and (2.6b) assert that capacity constraints are satisfied on sets of arcs, and (2.7) simply demands that the flow through each arc be nonnegative. Our objective is to find a feasible flow which has maximum value, i.e., one which maximizes

$$v = f(A_s) - f(B_s) = f(B_t) - f(A_t)\,.$$

Comment. It is possible to generalize the polymatroidal network flow model to provide for costs on arc flows and lower bounds on arc flows as imposed by supermodular set functions. However these generalizations are unnecessary for our present purposes.

3. Integrality and max-flow min-cut theorems

We now state two theorems proved in [10]. Strictly speaking, the statements of these theorems should allow for the possibility of the nonexistence of a maximal flow (which occurs if the maximum flow value is unbounded). However, we ignore this possibility, since it does not occur in any of the problems considered here.

Integrality Theorem. *If all capacity functions are integer-valued, then there exists a maximal flow which is integral.*

An *arc-partitioned cut* (S, T, L, U) is defined by a partition of the nodes into two sets S and T, with $s \in S$ and $t \in T$, and by a partition of the forward arcs across the cut into two sets L and U. The capacity of an arc partitioned cut is defined as

$$c(S, T, L, U) = \sum_{i \in S} \alpha_i(U \cap A_i) + \sum_{j \in T} \beta_j(L \cap B_j) .$$

As in the case of ordinary flow networks, the value v of any feasible flow f is equal to the net flow across any cut, i.e.,

$$v = f(L) + f(U) - f(B) ,$$

where B is the set of backward arcs, and clearly

$$v \leq c(S, T, L, U) .$$

Max-Flow Min-Cut Theorem. *The maximum value of a flow is equal to the minimum capacity of an arc-partitioned cut.*

4. (Poly)matroid intersection

The (unweighted) *matroid intersection* problem is as follows: Given two matroids (E, ρ_1) and (E, ρ_2), find a largest set $I \subseteq E$ such that I is independent in each of the matroids.

This problem can be formulated and solved as a flow problem as shown in Fig. 1. There are two nodes, s and t, and each arc from s to t corresponds to an element $e_i \in E$. The two capacity functions are determined by the two matroid

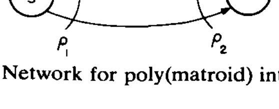


Fig. 1. Network for poly(matroid) intersection.

rank functions: $\alpha_s = \rho_1$, $\beta_t = \rho_2$. Since these capacity functions are integer-valued, there exists a maximal flow which is integral. Any such integral maximal flow corresponds to a solution to the matroid intersection problem.

Any partitioned cut (S, T, L, U) must have $S = \{s\}$, $T = \{t\}$ and is thus determined simply by a partition of E into two subsets L and $U = E - L$. The duality theorem for this problem thus follows as an immediate corollary of the max-flow min-cut theorem.

Matroid Intersection Duality Theorem. $\max|I| = \min_{L\subseteq E}\{\rho_1(E-L) + \rho_2(L)\}$.

In the *polymatroid intersection* problem, the capacity functions are simply the rank functions of the two polymatroids, and the duality theorem generalizes in an obvious way.

5. Matroid partitioning

A general version of the *matroid partitioning problem* is as follows: Given k matroids $M_j = (E, \rho_j)$, $j = 1, \ldots, k$, find a largest set $I \subseteq E$ such that I can be partitioned into k sets $I_1, \ldots, I_k$, where I_j is independent in (E, ρ_j).

A flow network for this problem is shown in Fig. 2. There is a node for each element $e_i \in E$, $i = 1, \ldots, n$, and a node for each matroid M_j plus a source s and a sink t. There is an arc (e_i, M_j) for each e_i and M_j. The flow into each node M_j is constrained by the capacity function $\beta_j = \rho_j$. Each arc (s, e_i) has unit capacity; for convenience we assume that these unit capacities are imposed by the capacity function α_s, where $\alpha_s(X) = |X|$, for all $X \subseteq A_s$. There are no other capacity constraints.

It follows from the integrality theorem that there exists a maximal flow which corresponds to an optimal solution to the partitioning problem. In order to determine an optimal dual solution we reason as follows. An arc-partitioned cut with finite capacity must have all nodes M_j in T. Let us denote by A the subset of nodes e_i in S; the nodes in $E - A$ are in T. For given S and T, L and

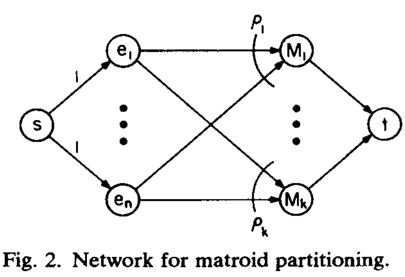


Fig. 2. Network for matroid partitioning.

U must be chosen as follows, in order for the arc-partitioned cut to have finite capacity:

$$L = \{(e_i, M_j) \mid e_i \in A\}, \qquad U = \{(s, e_i) \mid e_i \in E - A\},$$

as shown in Fig. 3. Thus the minimum capacity of an arc-partitioned cut is

$$\min_{A \subseteq E} \left\{ |E - A| + \sum_j \rho_j(A) \right\}.$$

This capacity is strictly less than $|E|$ if and only if

$$|A| > \sum_j \rho_j(A),$$

for some $A \subseteq E$. We have thus proved the following well-known result.

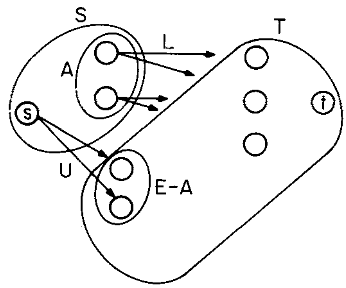


Fig. 3. Cut in proof of the matroid partitioning duality theorem.

Matroid Partitioning Duality Theorem (Edmonds and Fulkerson [2]). *Let I be a feasible solution to the matroid partitioning problem. Then*

$$\max |I| = \min_{A \subseteq E} \left\{ |E - A| + \sum_j \rho_j(A) \right\}.$$

Moreover, E is a feasible solution if and only if

$$|A| \leqslant \sum_j \rho_j(A),$$

for all $A \subseteq E$.

6. A problem of Krogdahl

The unweighted version of a problem considered by Krogdahl [9] and Schrijver [13] is as follows: Given k matroids $M_j = (E, \rho_j)$, $j = 1, \ldots, k$, and l matroids $M'_j = (E, \rho'_j)$, $j = 1, \ldots, l$, find a largest set I such that I can be partitioned into k subsets $I_1, \ldots, I_k$, where I_j is independent in M_j, and I can also be partitioned into l subsets $I'_1, \ldots, I'_l$, where I'_j is independent in M'_j.

It is well known that a set I can be so partitioned if and only if I is independent in both M and M', where M (M') is the *sum* of matroids $M_1, \ldots, M_k$ $(M'_1, \ldots, M'_l$, respectively). Thus this is actually a matroid intersection problem in which independence in each of the matroids M, M' can be determined by solving a matroid partitioning problem. Hence it is not surprising that the flow network for this problem, as shown in Fig. 4, is much like two networks for the matroid partitioning problem joined together. Each element e_i is represented by two nodes, e_i and e'_i, and an arc (e_i, e'_i), as shown in the center of the network. Each arc (e_i, e'_i) has unit capacity and we assume this capacity is imposed by a capacity function at the tail of the arc.

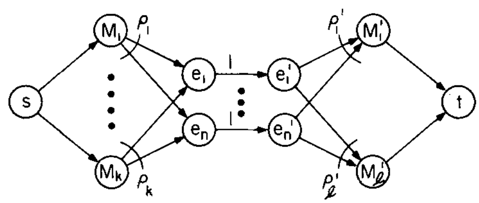


Fig. 4. Network for Krogdahl's problem.

In order for an arc-partitioned cut to have finite capacity, all of the nodes $M_1, \ldots, M_k$ must be in S and all of the nodes $M'_1, \ldots, M'_l$ must be in T. Thus we need only consider how the nodes $e_1, \ldots, e_n, e'_1, \ldots, e'_n$ are assigned to S and T:

(i) If $e'_i \in S$, then each of the arcs (e'_i, M'_j) must be in L. (There are no capacity constraints at the tails of these arcs.)

(ii) If $e_i \in T$, then each of the arcs (M_j, e_i) must be in U. (There are no capacity constraints at the heads of these arcs.)

(iii) If $e_i \in S$, $e'_i \in T$, then the arc (e_i, e'_i) must be in U, and its contribution to the capacity of the arc-partitioned cut is unity, independent of the other arcs in the cut.

It follows that there exists an arc-partitioned cut (S, T, L, U) of minimum capacity in which, for no arc (e_i, e'_i), the case holds that $e_i \in T$, $e'_i \in S$. If this were so, either e_i could be moved to S or e'_i could be moved to T without increasing the capacity. (Either the new L or the new U would be a proper subset of the old.) We thus see that an arc-partitioned cut simply partitions the arcs (e_i, e'_i) into three classes: A', A, and $E-(A \cup A')$, corresponding to the cases $e'_i \in S$, $e_i \in T$ and $e_i \in S$, $e'_i \in T$, indicated above, and as shown in Fig. 5.

By reasoning similar to that in the previous section, we now have the following result.

Theorem 6.1. *Let I be a feasible solution to the problem of Krogdahl. Then*

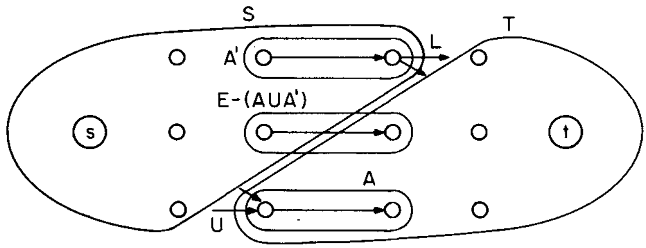


Fig. 5. Cut in proof of Theorem 6.1.

$$\max|I| = \min_{A,A' \subseteq E} \left\{ |E-(A \cup A')| + \sum_j \rho_j(A) + \sum_j \rho'_j(A') \right\}.$$

Moreover, E is a feasible solution if and only if for all A, A' ⊆ E,

$$|A \cup A'| \leq \sum_j \rho_j(A) + \sum_j \rho'_j(A').$$

A special case of interest is that of finding a largest common partial transversal of two families of subsets of E, $\mathscr{E} = (E_1, E_2, \ldots, E_k)$, and $\mathscr{E}' = (E'_1, \ldots, E'_l)$. Here let

$$\rho_j(X) = \begin{cases} 1 & \text{if } X \subseteq E_j, \\ 0 & \text{otherwise}, \end{cases}$$

and define ρ'_j similarly. Then the appropriate duality theorem follows.

As another special case, let $l = 1$ and k be arbitrary. Then the problem becomes that of finding a largest set I'_1, such that I'_1 is independent in M'_1 and I'_1 is partitionable into subsets $I_1, \ldots, I_l$, where I_j is independent in M_j. Since

$$\rho'_1(E-A) \leq |E-(A \cup A')| + \rho'_1(A'),$$

we have

$$\max|I'_1| = \min_{A \subseteq E} \left\{ \rho'_1(E-A) + \sum_j \rho_j(A) \right\}.$$

If $\rho'_1(X) = |X|$, for all $X \subseteq E$, then this result further specializes to the Edmonds–Fulkerson theorem.

7. A problem of Fujishige

Fujishige [5] has formulated and solved the following problem: Let $G = (V, E)$ be a digraph with two disjoint sets $V_1, V_2 \subseteq V$ of sources and sinks, respectively. Let (V_1, ρ_1) and (V_2, ρ_2) be polymatroids, and let each arc $(i, j) \in E$ be assigned a capacity c_{ij}. The problem is to find a maximum value flow from the sources in V_1 to the sinks in V_2 subject to the constraints that the net flows out of the source nodes in V_1 and the net flows into the sink nodes in V_2 are feasible with respect to the polymatroids (V_1, ρ_1), (V_2, ρ_2), and all arc capacity constraints are respected.

This problem can easily be formulated as a polymatroidal network flow problem by adding a dummy source s and a dummy sink t, as shown in Fig. 6.

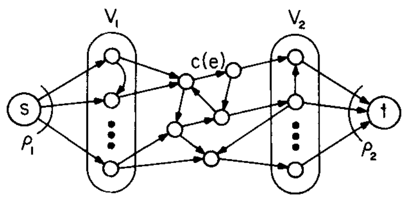


Fig. 6. Network for a problem of Fujishige.

There is an arc (s, i) to each node i in V_1 and we let $\alpha_s = \rho_1$. Similarly, there is an arc (j, t) from each node j in V_2 to t and we let $\beta_t = \rho_2$. We impose capacity constraints on the arcs (i, j) in G by letting

$$\alpha_i(X) = \sum_{(i,j)\in X} c_{ij} \quad (X \subseteq A_i).$$

An arc-partitioned cut for this network is indicated in Fig. 7. The forward arcs across this cut are of three kinds:

(i) Arcs of the form (s, i) where $i \in V_1 \cap T$. All such arcs are in U, if the cut is to have finite capacity, and their contribution to the capacity of the cut is $\rho_1(V_1 \cap T)$.

(ii) Arcs of the form (j, t), where $j \in V_2 \cap S$. All such arcs are in L, if the cut is to have finite capacity, and their contribution to the capacity of the cut is $\rho_2(V_2 \cap S)$.

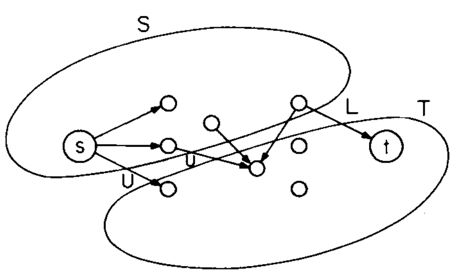


Fig. 7. Cut in proof of Theorem 7.1.

(iii) Each arc $e = (i, j)$ of G, such that $i \in S$, $j \in T$. The total contribution of these arcs to the capacity of the cut is

$$\sum_{i \in S, j \in T} c_{ij},$$

independent of the contributions of the arcs considered in (i) and (ii) above.

We thus have the following duality result.

Theorem 7.1 (Fujishige [4]). *Let v be the maximum feasible flow value. Then*

$$v = \min\left\{\rho_1(V_1 \cap S) + \rho_2(V_2 \cap T) + \sum_{i \in S, j \in T} c_{ij}\right\},$$

where the minimum is taken over all partitions S, T of V.

8. Gammoids and linking systems

Problems involving gammoids and linking systems are easily formulated as network flow problems. One such problem considered by Brualdi [1] and Schrijver [13] is as follows: Let $G = (V, E)$ be a digraph, V_1, V_2 be disjoint subsets of V, and $M_1 = (V_1, \rho_1)$, $M_2 = (V_2, \rho_2)$ be matroids. Find a pair of subsets $I_1 \subseteq V_1$, $I_2 \subseteq V_2$, such that $k = |I_1| = |I_2|$ is as large as possible, I_j is independent in M_j, $j = 1, 2$, and there are k node-disjoint paths linking I_1 to I_2.

This problem is clearly very much like that of Fujishige, except that, instead of arc capacities, there are node capacities (i.e., there may be at most one unit of flow through each node $i \in V$). In order to impose these constraints, we set

$$\alpha_i(X) = 1 \quad (\emptyset \neq X \subseteq A_i)$$

for each node $i \in V$. The following result is now easy to obtain, by the same reasoning as in the previous section.

Theorem 8.1 (Brualdi [1]). *The maximum of $k = |I_1| = |I_2|$ is equal to*

$$\min\{\rho_1(V_1 - I_1) + \rho_2(V_2 - I_2) + |X|\},$$

where $X \subseteq V$ intersects each path from I_1 to I_2.

9. Independent assignments

The unweighted version of the 'independent assignment' problem considered by Iri and Tomizawa [8] is as follows: Given a bipartite graph $G = (V_1, V_2; E)$ and two matroids $M_1 = (V_1, \rho_1)$, $M_2 = (V_2, \rho_2)$, find a matching in G such that the vertices $I_1 \subseteq V_1$, $I_2 \subseteq V_2$ covered by the matching are independent in M_1 and M_2.

This problem, which can also be formulated as a matroid intersection problem, is clearly a special case of the problem of Fujishige, where $V = V_1 \cup V_2$, and $c_{ij} = \infty$ for each arc $(i, j) \in E$, $i \in V_1$, $j \in V_2$. In order for an arc-partitioned cut (S, T, L, U) to have finite capacity, there can be no arcs (i, j) such that $i \in S$, $j \in T$. Thus the vertices $V_1' \cup V_2'$, where $V_1' = T \cap V_1$, $V_2' = S \cap V_2$, must provide a covering of the arcs of G. Define the *rank* of such a covering $V_1' \cup V_2'$ as $\rho_1(V_1') + \rho_2(V_2')$. We then have the following result.

Theorem 9.1. *The maximum size of an 'independent assignment' is equal to the minimum rank of a covering of edges by vertices.*

The well-known König–Egervary theorem is clearly a corollary, for the case that (V_j, ρ_j), $j = 1, 2$, are free matroids, i.e., $\rho_j(X) = |X|$, for all $X \subseteq V_j$.

10. The Rado–Hall problem

Let $M = (E, \rho)$ be a matroid and let $\mathscr{E} = (E_1, \ldots, E_k)$ be a family of subsets of E (with repetitions allowed). The problem is to find a largest partial transversal which is independent in M.

This is a matroid intersection problem, in which the two matroids are M and the transversal matroid induced by $\mathscr{E}$. The problem can also be viewed as a special case of the independent assignment problem in which the graph $G = (V_1, V_2; A)$ and the two matroids $M_1 = (V_1, \rho_1)$, $M_2 = (V_2, \rho_2)$ are as follows: $V_1 = E$, $V_2 = \mathscr{E}$, $(i, j) \in A$ if and only if $e_i \in E_j$, $\rho_1 = \rho$, $\rho_2(V_2') = |V_2'|$ for all $V_2' \subseteq V_2$. It is now easy to obtain the following result.

Rado–Hall Theorem (Rado [12]). *The size of a largest independent partial transversal is*

$$\min_{\mathscr{E}' \subseteq \mathscr{E}} \{|\mathscr{E} - \mathscr{E}'| + \rho(E')\},$$

where

$$E' = \bigcup_{E_j \in \mathscr{E}'} E_j .$$

Moreover, there exists an independent transversal if and only if the union of any k sets in $\mathscr{E}$ has matroid rank at least k.

Since this theorem specializes to the well-known Hall theorem in the case that M is a free matroid, it is known as the Rado–Hall theorem.

11. Concluding remarks

We believe that the formulations given in this paper demonstrate the usefulness of the polymatroidal network flow model. It should be emphasized that other models, e.g., those of Fujishige [5] and Edmonds and Giles [3], are equivalent in the sense that any problem which can be formulated and solved in terms of one model can be formulated and solved in terms of one of the others. (We assert, but do not prove here, that the Edmonds–Giles model can be reformulated in terms of the polymatroidal network flow model, with costs on the arcs.) We believe that the advantage of our model is that it permits problem formulations to be particularly simple and transparent. For example, although one can, in principle, reformulate the matroid partition problem as a matroid intersection problem, it seems simpler and more direct to formulate it directly in terms of the network given in Section 5.

References

[1] R.A. Brualdi, Menger's theorem and matroids, J. London Math. Soc. 4 (1971) 46–50.
[2] J. Edmonds and D.R. Fulkerson, Transversals and matroid partition, J. Res. Nat. Bur. Standards B69 (1965) 147–153.
[3] J. Edmonds and R. Giles, A min–max relation for submodular functions on graphs, Ann. Discrete Math. 1 (1977) 185–204.
[4] S. Fujishige, An algorithm for finding an optimal independent linkage, J. Oper. Res. Soc. Japan 20 (1977) 59–75.
[5] S. Fujishige, Algorithms for solving the independent flow problems, J. Oper. Res. Soc. Japan 21 (1978) 189–204.
[6] A. Frank, An algorithm for submodular functions on graphs, preprint, 1981.
[7] R. Hassin, On network flows, Ph.D. dissertation, Yale University, 1978.
[8] M. Iri and N. Tomizawa, An algorithm for finding an optimal independent assignment, J. Oper. Res. Soc. Japan 19 (1976) 32–57.
[9] S. Krogdahl, A combinatorial proof for a weighted matroid intersection algorithm, Univ. of Tromsø, Comp. Sci. Rept. 17, Tromsø, 1976.
[10] E.L. Lawler and C.U. Martel, Computing maximal 'polymatroidal' network flows, Math. Oper. Res., to appear.
[11] L. Lovász, The matroid matching problem, in: Algebraic Methods in Graph Theory, Colloq. Math. Soc. János Bolyai 25 (1978) 495–517.
[12] R. Rado, A theorem on independence relations, Quart. J. Math. (Oxford) 13 (1942) 83–89.
[13] A. Schrijver, Matroids and linking systems, Math. Centrum Tract 88, Math. Centrum, Amsterdam, 1978.
[14] P. Schönsleben, Ganzzahlige Polymatroid-Intersektions-Algorithmen, Thesis, ETH Zürich, 1980 (in German).
[15] U. Zimmermann, Minimization of some nonlinear functions over polymatroidal flows, preprint, 1981.

Annals of Discrete Mathematics 16 (1982) 201–211

TWO LINES LEAST SQUARES

A.K. LENSTRA and J.K. LENSTRA
Mathematical Centre, Amsterdam, The Netherlands

A.H.G. RINNOOY KAN
Econometric Institute, Erasmus University, Rotterdam, The Netherlands

T.J. WANSBEEK
Netherlands Central Bureau of Statistics, Voorburg, The Netherlands

It is well known that the standard single line least squares problem for n points in the plane is solvable in linear time. We consider two generalizations of this problem, in which two lines have to be constructed in such a way that, after a certain assignment of each point to one of the lines, the sum of squared vertical distances is minimal. Polynomial time algorithms for the solution of these problems are presented.

1. Introduction

Given a set $P = \{(x_j, y_j) \mid x_j, y_j \in \mathbb{R}, j = 1, \ldots, n\}$ of n points in the plane, the *single line least squares* (1LLS) problem is to find a line

$$l(x) = ax + b$$

such that

$$\sum_{(x,y)\in P} (l(x) - y)^2$$

is minimized. As is well known, the solution is given by

$$a = \frac{n \sum_P xy - \sum_P x \sum_P y}{n \sum_P x^2 - \left(\sum_P x\right)^2}, \qquad b = \frac{1}{n}\left(\sum_P y - a \sum_P x\right),$$

and the line l can thus be determined in $O(n)$ time.

In this paper we shall study two variations on this problem, in which *two lines* have to be constructed in such a way that, after a certain assignment of each point to one of the lines, the sum of the squared vertical distances is minimal.

The first and most obvious variation, the *two lines least squares* (2LLS) problem, is to find a set $Q \subseteq P$ and two lines

$$l_1(x) = a_1 x + b_1 , \qquad l_2(x) = a_2 x + b_2$$

such that

$$\sum_{(x,y)\in Q} (l_1(x) - y)^2 + \sum_{(x,y)\in \bar{Q}} (l_2(x) - y)^2 \tag{1}$$

is minimized (cf. Fig. 1).

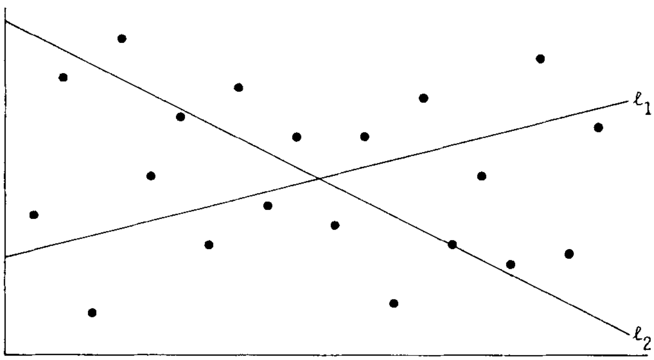


Fig. 1. 2LLS.

Another variation, the *bent line least squares* (BLLS) problem, is to find a *breakpoint* x^* and a *bent line*

$$l^*(x) = \begin{cases} a_1(x - x^*) + b & (x \leqslant x^*), \\ a_2(x - x^*) + b & (x > x^*) \end{cases}$$

such that

$$\sum_{(x,y)\in P} (l^*(x) - y)^2 \tag{2}$$

is minimized (cf. Fig. 2). Note that the bent line l^* is continuous in the breakpoint x^*.

These types of problems may well arise in many situations in applied statistics, e.g., biometrics and econometrics. If the observations belong to either

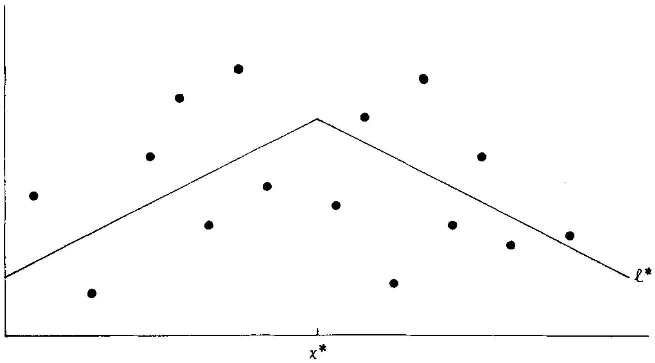


Fig. 2. BLLS.

of two regression regimes, with little or no classifying information being available, a two lines or bent line model may be more appropriate than the single line one.

The 2LLS problem occurs for example in the case of markets in disequilibrium, where data are available on prices and supply or demand, while it is not known whether each particular observation is generated by the supply curve or by the demand curve. This problem was considered by Fair and Jaffee [3], who suggested to obtain a solution by exhaustive search over all sets $Q \subseteq P$; this approach requires exponential time. In Section 2 we develop a polynomial algorithm to minimize (1) in $O(n^3)$ time.

The BLLS problem arises when, in the above situation, the observations correspond to the minimum of supply and demand. This is an example of a class of problems where the regression parameters may change as an independent variable increases. In this area an extensive literature has appeared; see [13] for a bibliography. The seminal paper on the BLLS problem is by Hudson [10], who also considered several generalizations involving multiple breakpoints. In Section 3 we use his results to minimize (2) in $O(n \log n)$ time.

Finally, in Section 4 we comment on the statistical properties of our estimators, on related previous work, and on possible extensions of our algorithms.

2. The two lines least squares problem

To solve the 2LLS problem, we start by observing two obvious properties of an optimal solution. First, the lines l_1 and l_2 are the ordinary 1LLS solutions for the sets Q and $\bar{Q}$ respectively. Secondly, the set $Q \subseteq P$ evidently satisfies

$$Q = \{(x, y) \mid |l_1(x) - y| \leq |l_2(x) - y|\}. \tag{3}$$

The partition of P into Q and $\bar{Q}$, given the lines l_1 and l_2, is therefore characterized by the set $Q_0 \subset \bar{\mathbb{R}}^2$ of points for which equality holds in (3):

$$\begin{aligned} Q_0 &= \{(x, y) \mid |l_1(x) - y| = |l_2(x) - y|\} \\ &= \{(x, y) \mid l_1(x) - y = l_2(x) - y\} \cup \{(x, y) \mid l_1(x) - y = -l_2(x) + y\} \\ &= \{(x_0, y)\} \cup \{(x, l_0(x))\} \end{aligned}$$

where

$$x_0 = \begin{cases} \dfrac{b_2 - b_1}{a_1 - a_2} & \text{if } a_1 \neq a_2\,, \\ +\infty & \text{if } a_1 = a_2,\ b_1 < b_2\,, \\ -\infty & \text{if } a_1 = a_2,\ b_1 \geq b_2\,, \end{cases}$$

$$l_0(x) = \tfrac{1}{2}(l_1(x) + l_2(x)).$$

Thus, the set Q_0 consists of two lines: the vertical line through the x-coordinate x_0 of the intersection of l_1 and l_2, and the median l_0 of l_1 and l_2. Under the assumption that $a_1 \geq a_2$, the set Q can now be rewritten as

$$Q = \{(x, y) \mid x \leq x_0,\ y \leq l_0(x)\} \cup \{(x, y) \mid x \geq x_0,\ y \geq l_0(x)\} \tag{4}$$

(cf. Fig. 3).

We conclude that we may restrict our attention to feasible solutions for which the partition of P into Q and $\bar{Q}$ is defined by a value x_0 and a line l_0 as in (4), and that a solution that is optimal with respect to such a partition is given by the 1LLS solutions for Q and $\bar{Q}$. (Note, however, that solutions satisfying (4) do not necessarily satisfy (3).) It follows that the 2LLS problem can be solved by generating all partitions of the above type, by solving two

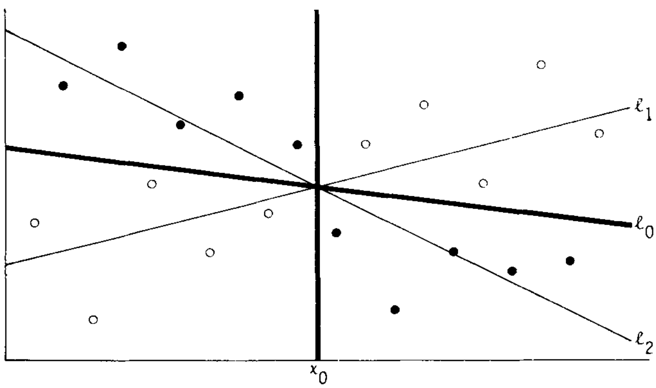


Fig. 3. The sets Q_0 (the heavy lines) and Q (the open points).

1LLS problems for each of them, and by selecting a solution for which the optimality criterion has minimal value.

What is the total number of different partitions of P of the type characterized by (4)? We may assume (if necessary, after a small perturbation) that no two points from P have the same x-coordinate and that no three points from P lie on the same line. First, it is clear that P can be separated into two subsets by a vertical line in n different ways, corresponding to the choices $x_0 = x_j$ $(j = 1, \ldots, n)$. Secondly, we claim that there is a one-to-one correspondence between separations of P by an arbitrary line and pairs of points from P, where the latter can be chosen in $\frac{1}{2}n(n-1)$ different ways. It follows that we have to consider no more than $\frac{1}{2}n^2(n-1)$ different partitions of P.

To see why the above claim is true, consider a separation of P into Q and $\bar{Q}$ by an arbitrary line l. Let $C(Q)$ and $C(\bar{Q})$ denote the convex hulls of Q and $\bar{Q}$ respectively. Since l also separates $C(Q)$ and $C(\bar{Q})$, there exists a unique line l_0 such that

(a) $y \geqslant l_0(x)$ for all points (x, y) in one of the convex hulls, say $C(Q)$,

(b) $y \leqslant l_0(x)$ for all points (x, y) in the other convex hull $C(\bar{Q})$,

(c) there are two points $p' = (x', y') \in C(Q)$ and $\bar{p}' = (\bar{x}', \bar{y}') \in C(\bar{Q})$ with $x' > \bar{x}'$ such that $y' = l_0(x')$ and $\bar{y}' = l_0(\bar{x}')$

(cf. Fig. 4); l_0 is obtained from l by turning l counterclockwise until l is tangent to both $C(Q)$ and $C(\bar{Q})$. Moreover, the assumption that no three points from P lie on the same line implies that $p' \in Q$ and $\bar{p}' \in \bar{Q}$.

This establishes one part of the correspondence. Conversely, consider two points $p'' = (x'', y'') \in P$ and $\bar{p}'' = (\bar{x}'', \bar{y}'') \in P$ with $x'' > \bar{x}''$. Let l_0 denote the line through p'' and $\bar{p}''$. A separation of P into Q and $\bar{Q}$ is now obtained by defining

$$
\begin{aligned}
Q &= \{(x, y) \mid y > l_0(x)\} \cup \{(x'', y'')\}, \\
\bar{Q} &= \{(x, y) \mid y < l_0(x)\} \cup \{(\bar{x}'', \bar{y}'')\};
\end{aligned} \tag{5}
$$

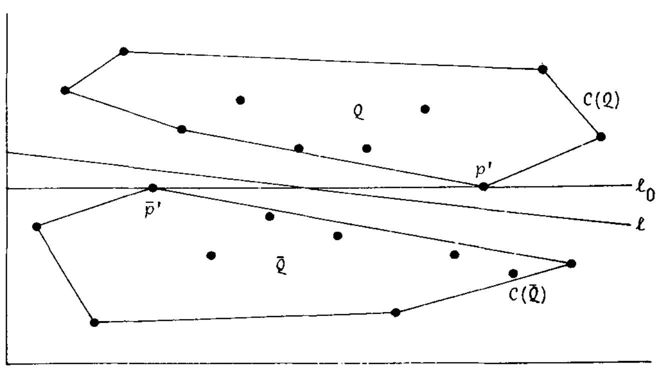


Fig. 4.

with respect to this separation, l_0 satisfies the conditions (a), (b) and (c) with $p' = p''$ and $\bar{p}' = \bar{p}''$.

If we start from a separation of P by an arbitrary line, apply the first transformation to find a pair of points and next apply the second transformation to this pair, we obtain the same separation once again. Thus, the correspondence is one-to-one, as claimed.

Our 2LLS algorithm can now be described as follows.

First, we renumber the points from P according to increasing x-coordinate in $O(n\log n)$ time.

Secondly, we consider all pairs of points $\{(x'', y''), (\bar{x}'', \bar{y}'')\}$ with $x'' > \bar{x}''$ in succession. For each such pair, we start by determining the sets Q and $\bar{Q}$ as in (5), calculating the partial sums

$$\sum_Q x, \quad \sum_Q y, \quad \sum_Q x^2, \quad \sum_Q xy, \tag{6}$$

and solving the 1LLS problem for Q and $\bar{Q}$; next, for $j = 1, \ldots, n-1$, we repeat this procedure for the partition induced by the vertical separation corresponding to $x_0 = x_j$. The sets Q and $\bar{Q}$ and the partial sums can be determined according to (5) and (6) in $O(n)$ time, and they can be adjusted for each successive value of x_0 in constant time. Given the partial sums (6), each 1LLS problem is solvable in constant time. It follows that this step requires $O(n)$ time for each pair of points, and $O(n^3)$ time overall.

Finally, we select the solution for which the optimality criterion achieves its minimal value. The entire procedure requires $O(n^3)$ time.

There are various ways in which the implementation of this algorithm can be improved that do not, however, reduce the running time by more than a constant factor. We note that, in general, it is impossible to generate all pairs of points from P in such a way that the partial sums (6) for a given separation can be derived from those for the preceding separation by interchanging a single point.

3. The bent line least squares problem

To solve the BLLS problem, we first renumber the points from P according to nondecreasing x-coordinate in $O(n\log n)$ time. We now reformulate the problem as follows: determine an index $k \in \{1, \ldots, n-1\}$ and two lines

$$l_1(x) = a_1 x + b_1, \qquad l_2(x) = a_2 x + b_2,$$

subject to a constraint on the x-coordinate of their intersection:

$$x_k \leqslant \frac{b_2 - b_1}{a_1 - a_2} \leqslant x_{k+1} \quad \text{if } a_1 \neq a_2\,, \tag{7}$$

$$b_1 = b_2 \qquad\qquad \text{if } a_1 = a_2\,,$$

such that

$$\sum_{j=1}^{k} (l_1(x_j) - y_j)^2 + \sum_{j=k+1}^{n} (l_2(x_j) - y_j)^2$$

is minimized. Since the breakpoint x^* has to be located in one of the intervals $[x_k, x_{k+1}]$, both formulations are clearly equivalent.

We shall show how, after an $O(n)$ initialization, the determination of l_1 and l_2 for any given value of k can be carried out in constant time. To select the optimal value of k, this has to be done for $k = 1, \ldots, n-1$, and the entire procedure requires $O(n\log n)$ time, as announced. Note that, apart from sorting the set P, our BLLS algorithm requires linear time off-line, whereas the 1LLS problem is solvable in linear time on-line.

We start by calculating the partial sums

$$\sum_{j=1}^{k} x_j\,, \qquad \sum_{j=1}^{k} y_j\,, \qquad \sum_{j=1}^{k} x_j^2\,, \qquad \sum_{j=1}^{k} x_j y_j \tag{8}$$

for $k = 1, \ldots, n$ in $O(n)$ time. Next, for a given value of k, we solve the ordinary 1LLS problem for $\{(x_j, y_j) \mid j = 1, \ldots, k\}$ and for $\{(x_j, y_j) \mid j = k+1, \ldots, n\}$ to find two lines l'_1 and l'_2 respectively; in view of the availability of the partial sums (8), this requires constant time.

In the x-coordinate of the intersection of l'_1 and l'_2 lies in $[x_k, x_{k+1}]$ or if $l'_1 = l'_2$, the pair (l'_1, l'_2) defines a feasible and optimal solution with respect to the given value of k, and we are finished.

If this is not the case, we claim that the optimal pair of lines has its intersection on either x_k or x_{k+1}, i.e., one of the inequalities in (7) has to be satisfied as an equality (cf. [10]). To see why, compare the infeasible pair (l'_1, l'_2) to any feasible solution (l_1, l_2) for which both inequalities in (7) are strictly satisfied. Suppose that $l'_2 \neq l_2$ and that these lines intersect in a point p (cf. Fig. 5). Then any line l''_2 through p whose direction is between the directions of l'_2 and l_2 yields an improvement over l_2, since the 1LLS optimality criterion is quadratic and convex in the parameters a and b of a line $ax + b$ and since l''_2 is closer than l_2 to the optimal line l'_2. If moreover the x-coordinate of the intersection of l_1 and l''_2 lies in $[x_k, x_{k+1}]$, the pair (l_1, l''_2) defines a feasible solution that is better than (l_1, l_2), and the latter solution cannot be optimal. The argument is easily extended to the case that l'_2 and l_2 are parallel.

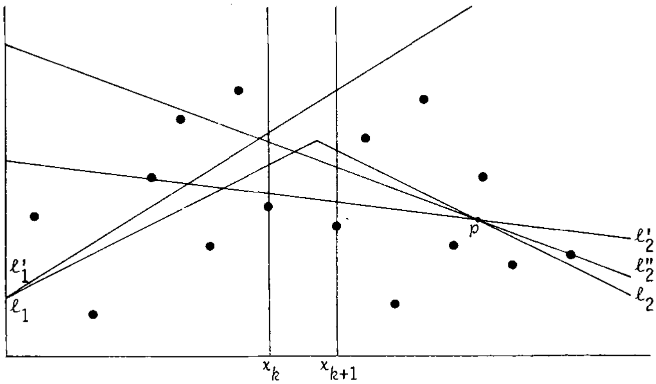


Fig. 5.

All we have to do in this case, therefore, is to determine two optimal bent lines l^* with fixed breakpoints $x^* = x_k$ and $x^* = x_{k+1}$ respectively and to select whichever of the two is best. The determination of l^* for a given value of x^* can be carried out in constant time, as follows. Starting from (2), we rewrite the optimality criterion as

$$\sum_{j=1}^{k} (a_1(x_j - x^*) + b - y_j)^2 + \sum_{j=k+1}^{n} (a_2(x_j - x^*) + b - y_j)^2 .$$

Taking first derivatives with respect to the parameters a_1, a_2 and b, we obtain the following linear equations:

$$\begin{cases} a_1 \sum_{j=1}^{k} (x_j - x^*)^2 + b \sum_{j=1}^{k} (x_j - x^*) = \sum_{j=1}^{k} (x_j - x^*) y_j , \\ a_2 \sum_{j=k+1}^{n} (x_j - x^*)^2 + b \sum_{j=k+1}^{n} (x_j - x^*) = \sum_{j=k+1}^{n} (x_j - x^*) y_j , \\ a_1 \sum_{j=1}^{k} (x_j - x^*) + a_2 \sum_{j=k+1}^{n} (x_j - x^*) + bn = \sum_{j=1}^{n} y_j . \end{cases}$$

Given the partial sums (8), this system is solvable in constant time to yield the required values of a_1, a_2 and b.

It is again an easy matter to conceive of various improvements in the implementation of this algorithm, when the above system has to be solved $O(n)$ times with only slight intermediate changes in the coefficients (cf. [10]). The worst case running time, however, will be affected by no more than a constant factor as a result.

4. Concluding remarks

The 2LLS and BLLS problems have so far been taken to be purely deterministic problems. Keeping in line with regression analysis tradition, we now assume that the parameters to be determined (Q, a_1, a_2, b_1, b_2 in 2LLS, x^*, a_1, a_2, b in BLLS) have true but unknown values that are to be estimated and raise the question what the statistical properties are of the outcomes of the minimization problems. With respect to the stochastic nature of the data generating process, we make the simple assumption that the x_j are non-stochastic and that the observations (x_j, y_j) satisfying the unknown linear relations are subject to vertical disturbances which are drawn from a normal distribution with zero expectation and constant variance. Under this interpretation, our *least squares* estimators are also *maximum likelihood* estimators.

The estimators in the 2LLS problem do probably not have any other desirable statistical properties. The following simple example shows that they are not in general consistent (which is not to say that consistent 2LLS estimators do not exist). Suppose that the true model is given by

$$y_1 = 1 + u\,, \qquad y_2 = -1 + u\,,$$

where u has a standard normal distribution with expectation 0, variance 1 and density function ϕ, i.e., $a_1 = a_2 = 0$, $b_1 = 1$, $b_2 = -1$. Suppose further that the independent variable x assumes at least three equidistant values, and that, when n goes to infinity, either regime accounts for half of the data with an equal number of data for each value of x. Then it is easy to see that the estimators $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_1$, $\hat{b}_2$ of a_1, a_2, b_1, b_2 satisfy

$$\operatorname{plim} \hat{a}_1 = a_1\,, \qquad \operatorname{plim} \hat{b}_1 = b_1 + c\,,$$
$$\operatorname{plim} \hat{a}_2 = a_2\,, \qquad \operatorname{plim} \hat{b}_2 = b_2 - c\,,$$

where

$$c = \int_1^\infty (2u-2)\phi(u)\,\mathrm{d}u \approx 0.167\,.$$

In this situation, the estimators of b_1 and b_2 are bounded away from the true values, due to persistent misclassification. (The assertion by Fair and Jaffee [3, p. 500] as to the consistency of the 2LLS estimators is incorrect. They neglect the implicit presence of dummy variables assigning observations to regimes. The fact that the number of such variables goes to infinity with n invalidates standard maximum likelihood theory.)

Statistical properties of the estimators in the BLLS problem follow from results due to Feder [4], who considered the case of multiple breakpoints and a more general functional specification for the regression segments. As the finite sample distribution of the estimators is intractable, he concentrated on their asymptotic distribution. His results imply that, if $a_1 \neq a_2$, the estimators of x^*, a_1, a_2, b are consistent and have a certain multinormal asymptotic distribution [4, pp. 71, 77]. Feder and Sylwester [5] already established the asymptotic normality of the estimator of x^*. Hinkley [9] showed that the asymptotic distribution is not a good approximation of the small sample distribution and presented an alternative that performs better in small samples.

Due to its combinatorial nature and the statistical intractability of its estimators, the 2LLS problem has not spawned much research. Mustonen [12] and Hermann [8] considered the multidimensional case, in which two hyperplanes [8] or, more generally, two functions of a given form [12] have to be constructed. In view of the nonlinearity of the optimality criterion, they suggested to obtain an approximate solution by iterative numerical methods. We conjecture that our combinatorial approach can be extended to yield an optimal solution in polynomial time, as long as the dimension and the number of hyperplanes are constants. However, this generalization is likely to be extremely intricate.

Another reason for the relative neglect of the 2LLS problem is that, as far as disequilibrium econometrics is concerned, economic theory can be invoked to further model the regime choice mechanism. Fair and Jaffee [3] made several additional assumptions, the simplest one being that a price increase points to an excess demand regime and a price fall to an excess supply regime. A huge literature has developed in this direction (e.g. [7, 2]), a common trait being the use of nonlinear rather than combinatorial optimization methods. Estimators with favorable asymptotic properties have been derived (e.g. [11]).

It has been mentioned that the BLLS problem in which multiple breakpoints are allowed has been considered by Hudson [10]. In contrast to his analytical approach, Bellman and Roth [1] proposed a dynamic programming recursion to obtain a solution which is approximate in the sense that the breakpoints are to be chosen on a grid. The running time of their method depends heavily on the grid width; it is linear in the number of breakpoints, but only pseudo-polynomial [6] in the data (x_j, y_j). For the many variations on and extensions of the BLLS problem, the reader is again referred to the bibliography in [13].

The purpose of this paper has been to analyze the computational complexity of two combinatorial optimization problems arising in statistical analysis. Other results of a similar nature can be found in the work of Shamos [14]. These examples should serve to demonstrate the potential value of research in this interface area.

References

[1] R. Bellman and R. Roth, Curve fitting by segmented straight lines, J. Amer. Statist. Assoc. 64 (1969) 1079–1084.

[2] R.J. Bowden, The Econometrics of Disequilibrium (North-Holland, Amsterdam, 1978).

[3] R.C. Fair and D.M. Jaffee, Methods of estimation for markets in disequilibrium, Econometrica 40 (1972) 497–514.

[4] P.I. Feder, On asymptotic distribution theory in segmented regression problems—identified case, Ann. Statist. 3 (1975) 49–83.

[5] P.I. Feder and D.L. Sylwester, On the asymptotic theory of least squares estimation in segmented regression: identified case (abstract), Ann. Math. Statist. 39 (1968) 1362.

[6] M.R. Garey and D.S. Johnson, 'Strong' NP-completeness results: motivation, examples and implications, J. Assoc. Comput. Mach. 25 (1978) 499–508.

[7] S.M. Goldfeld and R.E. Quandt, The estimation of structural shifts by switching regressions, Ann. Econ. Soc. Measurement 2 (1973) 475–485.

[8] J. Hermann, Data-analytic concepts for isolating intersecting reaction surfaces, Institute for Econometrics and Operations Research, University of Bonn, 1980.

[9] D.V. Hinkley, Inference about the intersection in two-phase regression, Biometrika 56 (1969) 495–504.

[10] D.J. Hudson, Fitting segmented curves whose join points have to be estimated, J. Amer. Statist. Assoc. 61 (1966) 1097–1129.

[11] N.M. Kiefer, Discrete parameter variation: efficient estimation of a switching regression model, Econometrica 46 (1978) 427–434.

[12] S. Mustonen, Digression analysis: fitting alternative regression models to heterogeneous data, in: L.C.A. Corsten and J. Hermans, eds., COMPSTAT '78: Proc. Computational Statistics (Physica-Verlag, Vienna, 1978) pp. 95–101.

[13] S.A. Shaban, Change point problem and two-phase regression: an annotated bibliography, Internat. Statist. Rev. 48 (1980) 83–93.

[14] M.I. Shamos, Geometry and statistics: problems at the interface, in: J.F. Traub, ed., Algorithms and Complexity: New Directions and Recent Results (Academic Press, New York, 1976) pp. 251–280.

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 213–223

BOUNDING THE INDEPENDENCE NUMBER OF A GRAPH

L. LOVÁSZ

Bolyai Institute, József Attila University, Szeged H-6720, Hungary

0. Introduction

The problem of computing the independence number of a graph is NP-complete; the matching number, on the other hand is computable in polynomial time. This difference in their computational complexity implies that, to attack these two problems, different strategies have to be applied. The matching problem may serve as a prototype of handling 'easy', i.e., polynomially solvable problems: find a good characterization, then a polynomial algorithm, describe the facets of associated polyhedra etc. None of these lines of attack holds out promises of success in the case of the independence number problem. So what scheme should one follow in the study of an NP-complete problem like independence number? Discouraged by the fact of NP-completeness, one might answer (or at least feel) that this problem is mathematically intractable and so one should not waste time on it. Others, who play down the relevance of polynomiality in algorithms, might say that all there is to do is to improve the (more or less trivial) exponential algorithms by heuristics, programming, data handling tricks, etc. These two extremes meet in the opinion that no further attempts to 'grip the essence' of the problem are needed.

In this paper we discuss an idea which *might* suggest some non-trivial approaches to NP-complete problems. Whether the sporadic phenomena collected here will ever fall into a pattern, and whether from this a unified approach to NP-complete problems can be learned, is beyond the guessing of this author.

Let $\alpha(G)$ denote the independence number of the graph G. If $\alpha(G)$ cannot be calculated efficiently, a next step is to ask for sharp bounds on it. There is a very significant difference between upper and lower bounds: A lower bound means to prove the *existence* of an independent set of some size, which is usually proved by more-or-less constructive methods (heuristics, random choice, etc.). On the other hand, an upper bound means the *non-existence* of an independent set of larger size—in this respect it is 'destructive'. Is there any practical value then in finding upper bounds? Potential applications of upper bounds are the following:

- In a branch-and-bound procedure, sharp upper bounds may cut down the size of the search tree considerably. (Question: is there any example in the case of an 'easy' problem where a polynomial-bounded algorithm is obtained by pruning a search tree with the help of an upper bound? Such an example might shed some light on the hitherto somewhat mysterious phenomenon that well-characterized problems tend to be polynomially solvable.)
- Given a reasonable upper bound, we may consider the class of those graphs for which this upper bound is attained, and then restrict the independence number problem to this class. There is a good chance that it is easier to solve that problem for this particular class of graphs, and such graphs may well be very interesting.
- Deriving sharper and sharper upper bounds, more and more insight could be gained into the nature of independence number (a procedure vaguely reminescent of the expansion of a function into, say, a Fourier series).

We shall survey some methods to obtain upper bounds on the independence number $\alpha(G)$ of a graph. We have left the precise notion of 'upper bound' open. The most natural choice, of course, is to look for a positive integer valued function φ defined on graphs, such that $\alpha(G) \leq \varphi(G)$ for every graph G and φ is polynomially computable. Sometimes we shall have to settle for less, e.g., the function φ should be such that $\varphi(G) \leq k$ is an NP-property of the pair (G, k). Putting things even more general, we shall be interested in methods which enable us to exhibit the relation $\varphi(G) \leq k$ for reasonably many pairs (G, k).

Let us remark that some of these results are easier to state in terms of $\tau(G) = |V(G)| - \alpha(G)$, the point-covering number of G. Also note that if $L(G)$ denotes the line-graph of G, then $\alpha(L(G))$ is the matching number of G, and so it is well-behaved. To what class of graphs generalizing line-graphs the successful theory of matchings can be extended is the motivation of some important current research [14, 16].

Finally, let us point out that complexity considerations concerning the independence number problem motivate, and may even initiate, research in fields like algebraic geometry, linear algebra and algebraic topology. Although these connections are in a very embryonic state, any link between graph theory (or combinatorial optimization) and these deep, classical fields of mathematics is, I feel, of particular interest.

1. Eigenvalues

We describe very briefly an upper bound on α which was discovered in connection with a problem of Shannon in coding theory.

Let G be a graph on $V(G) = \{1, \ldots, n\}$ and let $\mathscr{A}$ denote the set of $n \times n$ symmetric matrices $A = (a_{ij})$ such that $a_{ij} = 1$ if $i = j$, or i and j are non-adjacent. Let $\Lambda(A)$ denote the largest eigenvalue of A. Define

$$\vartheta(G) = \min\{\Lambda(A) \colon A \in \mathscr{A}\}.$$

Since every $A \in \mathscr{A}$ contains a symmetric $\alpha(G) \times \alpha(G)$ submatrix J of all 1's, it follows that

$$\Lambda(A) \geqslant \Lambda(J) = \alpha(G),$$

and so $\vartheta(G) \geqslant \alpha(G)$. What is important about ϑ is that it is polynomially computable. (More precisely, for every $\varepsilon > 0$ a rational approximation of $\vartheta(G)$ with error less than ε can be computed in time polynomial in $|\log \varepsilon|$ and n. Note that ϑ may be irrational !) The idea of computing $\vartheta(G)$ is that $\Lambda(A)$ is a convex function of A on the affine subspace $\mathscr{A}$, and it can be minimized using the methods of Shor [17] and Yudin and Nemirovskii [18] (see also [4]).

As remarked before, this function ϑ gives rise to a class of graphs for which α is efficiently computable, namely the class of graphs with $\alpha(G) = \vartheta(G)$. This class is, however, rather ugly: it is in NP but it is also NP-complete. A nicer subclass is the class of perfect graphs. For perfect graphs the only known polynomial algorithm to compute $\alpha(G)$ is through computing $\vartheta(G)$ (see [4]).

It was also remarked that if a polynomially computable upper bound is found, then this can be used to prune branch and bound search trees. Experience shows that the use of ϑ does prune the search for maximum independent set considerably [3], but no theoretical results have been obtained so far concerning the size of the 'pruned' tree.

2. Algebraic geometry

The section title is perhaps somewhat immodest, but the flavor of the result of Li and Li [9], which is the starting point of our discussion, is indeed algebraic geometry.

Let G be a simple graph on $V(G) = \{1, \ldots, n\}$, and let us consider n variables $x_1, \ldots, x_n$. Form the polynomial

$$f(G; x_1, \ldots, x_n) = \prod_{i,j \in E(G)} (x_i - x_j).$$

(This polynomial depends on the labelling of the points, but only up to its sign, which shall play no role.) Note the following simple fact.

Lemma 2.1. *$\alpha(G) \leq k$ iff, identifying $k+1$ variables in $f(G; x_1, \ldots, x_n)$ in all possible ways, we always obtain the zero polynomial.*

Let $X_k^n \subseteq C^n$ denote the set of those vectors which have at least $k+1$ equal coordinates, and let I_k^n denote the ideal of those polynomials in $C[x_1, \ldots, x_n]$ which vanish for every vector in X_k^n. Thus $\alpha(G) \leq k$ iff $f(G; x_1, \ldots, x_n) \in I_k^n$. To make use of this observation one needs a description of I_k^n which will enable us to exhibit 'easily' that a polynomial is in I_k^n. A natural approach is to find generators for I_k^n, and this was indeed accomplished by Li and Li [9]. Let $\mathcal{H}_k^n$ denote the set of those graphs on $\{1, \ldots, n\}$ which are unions of k disjoint complete graphs. Let $\bar{\mathcal{H}}_k^n$ denote the subset of $\mathcal{H}_k^n$ consisting of those graphs where the sizes of the components are as equal as possible (i.e., every graph in $\bar{\mathcal{H}}_k^n$ consist of $n - k\lfloor n/k \rfloor$ copies of a complete $\lceil n/k \rceil$-graph and $k\lceil n/k \rceil - n$ copies of a complete $\lfloor n/k \rfloor$-graph. Note that all members of $\bar{\mathcal{H}}_k^n$ are isomorphic, but we are considering labelled graphs !)

Theorem 2.2. *I_n^k is generated by the polynomials $f(H; x_1, \ldots, x_n)$ $(H \in \bar{\mathcal{H}}_k^n)$.*

Corollary 2.3. *A graph G satisfies $\alpha(G) \leq k$ iff there exist polynomials $g_H(x_1, \ldots, x_n)$ $(H \in \bar{\mathcal{H}}_k^n)$ such that*

$$f(G; x_1, \ldots, x_n) = \sum_{H \in \bar{\mathcal{H}}_k^n} g_H(x_1, \ldots, x_n) f(H; x_1, \ldots, x_n).$$

Of course, Theorem 2.2 and Corollary 2.3 remain true if $\bar{\mathcal{H}}_k^n$ is replaced by $\mathcal{H}_k^n$, and the main difficulty lies in proving these weaker conclusions.

Let us remark that if $(x_1, \ldots, x_n) \in C^n$ and $f(H; x_1, \ldots, x_n) = 0$ for every $H \in \bar{\mathcal{H}}_k^n$, then an easy argument shows that $(x_1, \ldots, x_n) \in X_k^n$. Hence, by the Nullstellensatz of Hilbert, there exist a natural number $p > 0$ and polynomials $g_H(x_1, \ldots, x_n)$ such that

$$f(G; x_1, \ldots, x_n)^p = \sum_{H \in \bar{\mathcal{H}}_k^n} g_H(x_1, \ldots, x_n) f(H; x_1, \ldots, x_n).$$

The main contents of Theorem 2.2 is that $p = 1$. This is somewhat reminiscent of the situation in integer linear programming, where a minimax formula for a linear relation follows in generality by the Duality Theorem, and one has to work hard and use special features of the problem to show that the denominators of the optimal solution of the linear program are 1's.

Why is this an upper bound on $\alpha(G)$? In the general sense mentioned in the introduction, $\alpha(G) \leq k$ can be proved by exhibiting polynomials g_H $(H \in \mathcal{H}_k^n)$ such that

$$(\mathrm{R}_1) \qquad f(G; x_1, \ldots, x_n) = \sum_{H \in \bar{\mathscr{H}}_k^n} g_H(x_1, \ldots, x_n) f(H; x_1, \ldots, x_n).$$

There are, however, three problems which arise here:

(1) The cardinality of $\bar{\mathscr{H}}_k^n$ is exponential in n; so (R_1) can be written down in polynomial time only if all but a polynomial number of the g_H are 0.

(2) It might happen that although (R_1) has only a polynomial number of terms, the coefficients g_H cannot be written down in polynomial space.

(3) Even if (R_1) is written down, there may not be any procedure to verify it in polynomial time.

Of these, the first problem is really serious and it limits the applicability of (R_1) to prove $\alpha(G) \leqslant k$ to special classes of graphs. (Question: can one *prove* that there exist graphs G for which every equation of type (R_1) has exponentially many terms on the right-hand side ?)

Objection (2) can be eliminated. I have proved that in (R_1) the coefficients g_H themselves may be chosen in the form $f(G_H; x_1, \ldots, x_n)$ with some graphs G_H. Thus the following theorem may be formulated.

Theorem 2.4. *A graph G has $\alpha(G) \leqslant k$ if and only if there exist graphs $G_1, \ldots, G_m$ on $V(G)$ such that each G_i can be partitioned into k cliques and*

$$(\mathrm{R}_2) \qquad f(G; x_1, \ldots, x_n) = \sum_{i=1}^{m} f(G_i; x_1, \ldots, x_n).$$

Finally, objection (3) is only moderately serious.

Of course, we cannot simply expand all polynomials occurring and then see if all terms cancel (as we learn at school), since the expansion of just $f(G; x_1, \ldots, x_n)$ contains exponentially many terms. But we may, say, generate values for $x_1, \ldots, x_n$ at random, substitute, and see if the two sides are equal. If they are not, then, of course, we know that (R_1) does not hold. If they are, then the probability that (R_1) is not an identity is negligible, but we have hit a choice of variables for which the two sides are equal. So the verification of (R_1) can be carried out at least in a 'random polynomial' framework. The problem of verifying a polynomial identity in deterministic polynomial time is an outstanding problem in the complexity theory of algebra. It may well be, however, that special identities like (R_2) can be verified easier.

Kleitman and the present author observed that a 'dual' version of the theorem of Li–Li is also true (in fact, it is easier to prove). Let $Y_k^n \subseteq C^n$ denote the set of those vectors which have at most k distinct entries, and let J_k^n denote the ideal of those polynomials which vanish for every vector in Y_k^n.

Lemma 2.5. *A graph G has chromatic number $\geq k$ if $f(G; x_1, \ldots, x_n) \in J_k^n$.*

The less trivial part is the following. Let $\mathscr{L}_k^n$ denote the set of those graphs on $\{1, \ldots, n\}$ whose edges form a complete k-graph (and which have, therefore, $n-k$ isolated points).

Theorem 2.6. *The polynomials $f(L; x_1, \ldots, x_n)$ $(L \in \mathscr{L}_k^n)$ generate the ideal J_k^n.*

Corollary 2.7. *A graph G satisfies $\chi(G) \geq k$ iff there exist polynomials $g_L(x_1, \ldots, x_n)$ $(L \in \mathscr{L}_k^n)$ such that*

$$f(G; x_1, \ldots, x_n) = \sum_{L \in \mathscr{L}_k^n} g_L(x_1, \ldots, x_n) f(L; x_1, \ldots, x_n).$$

Again, the following sharper version is true.

Theorem 2.8. *A graph G satisfies $\chi(G) \leq k$ iff there exist graphs $G_1, \ldots, G_m$ on $V(G)$, each containing a complete k-graph, such that*

$$f(G; x_1, \ldots, x_n) = \sum_{i=1}^{m} f(G_i, x_1, \ldots, x_n).$$

In this last form this result is reminiscent of a well-known result of Hajós [6], which also yields a 'pseudo-good' characterization of graphs with chromatic number $\geq k$. Define 3 operations on the set of graphs:

(α) add new points and/or lines,

(β) identify two non-adjacent points,

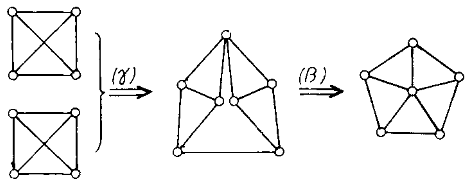
(γ) take two graphs G_1, G_2, delete two edges $x_1y_1 \in E(G_1)$ and $x_2y_2 \in E(G_2)$, identify x_1 with x_2 and join y_1 to y_2 by a new edge.

Theorem 2.9. *A graph G has $\chi(G) \geq k$ iff it can be constructed from complete k-graphs by the repeated application of steps (α), (β) and (γ).*

Again, the relation $\chi(G) \geq k$ can be proved for a graph G by carrying out the construction explicitly. Just how short this proof is, depends on the graph G. Perfect graphs can be obtained in one step. Are there other interesting classes for which the construction is short ? So far, Hajós' theorem was studied for its possible applications to planarity; its algorithmic complexity aspects are an unexplored territory (cf. Fig. 1).

To this approach to chromatic number the same remarks apply as to the Li–Li theorem on independence number.

Li and Li point out that by looking at the degrees of the polynomials, one



$$(x_1-x_2)(x_2-x_3)(x_3-x_4)(x_4-x_5)(x_5-x_1)(x_1-x_6)(x_2-x_6)(x_3-x_6)(x_4-x_6)(x_5-x_6)=$$
$$=(x_1-x_2)(x_2-x_3)(x_1-x_3)(x_1-x_6)(x_2-x_6)(x_3-x_6)(x_3-x_4)(x_4-x_5)(x_4-x_6)(x_5-x_6)$$
$$+(x_1-x_2)(x_2-x_3)(x_1-x_6)(x_2-x_6)(x_3-x_4)(x_4-x_5)(x_3-x_5)(x_3-x_6)(x_4-x_6)(x_5-x_6)$$

Fig. 1. Proving that the 5-wheel has chromatic number $\geqslant 4$ by Hajós' construction and by the method of polynomials.

sees immediately that if $\alpha(G)\leqslant k$, then

$$|E(G)| = \deg f(G; x_1, \ldots, x_n) \geqslant \deg f(H; x_1, \ldots, x_n),$$

where $H \in \mathscr{H}_k^n$. This result is just Turán's theorem (for the complement of G). They also obtain generalizations of Turán's theorem this way, but we cannot go into the details of this.

The idea of using the degree of a polynomial to obtain combinatorial estimations also occurs in a paper by Brouwer and Schrijver [2], where they use it to calculate $\tau(H)$ for the hypergraph H formed by the lines of an affine plane over a finite field.

Let us conclude this section with the remark that Hilbert's Nullstellensatz may well be a source of other interesting 'good' or 'pseudo-good' characterizations in combinatorics. More generally, the duality between 'syntax' and 'semantics' comes up here (in the Nullstellensatz, the solvability of a system of algebraic equations—a 'semantical' problem—is characterized in terms of the non-expressability of 1 as an element of the ideal generated by the left-hand side—a syntactical property). So, e.g. Gödel's Completeness Theorem could be viewed as a 'pseudo-good' characterization of the consistency of a system of axioms: if it is inconsistent, we can exhibit this by deriving a contradiction, if it is consistent, we can exhibit a model. Of course, no polynomiality (or even finiteness) of these procedures is claimed. Whether polynomiality enters the picture in any reasonable way is not known.

3. Matroids

These results (see [10]) are best discussed in terms of the point-covering number $\tau(G)$. Let us assume that a matroid $(V(G), r)$ is introduced on $V(G)$. Then we may generalize the problem of determining $\tau(G)$ to determining $\tau(G, r)$, the minimum rank of a point-cover. In the special case where $(V(G), r)$ is the free matroid, we have $\tau(G, r) = \tau(G)$.

The gain in introducing this matroid structure on $V(G)$ is that we have more freedom in applying some reduction procedures. Let $v \in V(G)$, $r(v) = 1$ and assume that v is in the flat spanned by its neighbours. Delete v from the graph and contract v in the matroid. Then the resulting graph G' and matroid $(V(G'), r')$ satisfy

$$\tau(G', r') = \tau(G, r) - 1 . \tag{1}$$

If $v \in V(G)$ has rank $r(v) = 0$, then for the graph G' and rank function r' obtained similarly as above we have

$$\tau(G', r') = \tau(G, r) . \tag{2}$$

If $v \in V(G)$ is a coloop in the matroid, then let $(V(G), r'')$ be a new matroid which is obtained by deleting v, place a 'general' point on the flat spanned by its neighbours, and finally label this new point v.

Then

$$\tau(G, r'') = \tau(G, r) . \tag{3}$$

If $(V(G), r')$ is any weak map of the matroid $(V(G), r)$ (i.e., $r' \leqslant r$), then trivially

$$\tau(G, r') \leqslant \tau(G, r) . \tag{4}$$

Now these reductions enable us to prove the relation $\tau(G) \geqslant k$ for quite a few graphs. (It is not clear which are interesting classes for which this can be accomplished.) Fig. 2 shows how to exhibit, using reductions (1)–(4) that the graph G has $\tau(G) \geqslant 6$.

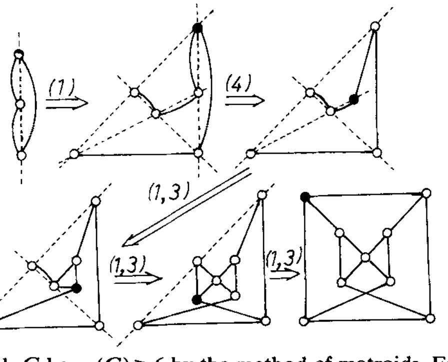


Fig. 2. Proving that the graph G has $\tau(G) \geqslant 6$ by the method of matroids. For brevity, (1) followed by (3) is depicted as one step, $(1, 3)$.

So far, the main application of this method has been to develop a classification theory for τ-critical graphs (Lovász [11]). In the algorithmic context, an important problem which arises is handling the matroids. A natural approach is to restrict oneself to real representable matroids, and then handle them as real matrices. There are, however, many problems in connection with this, for example, how to construct a representation of a principal extension from a representation of the matroid? We cannot go into the complicated problems arising here.

4. Topology

The present author [12] proved the following lower bound on the chromatic number $\chi(G)$ of a graph. Let us define the neighbourhood complex $\mathcal{N}(G)$ of a graph G as the simplicial complex whose vertices are the points of G and whose simplices are those subsets of $V(G)$ which have a neighbour in common. Let us recall that a topological space T is called *k-connected* if for every $0 \leq r \leq k$, every continuous map of the r-sphere S^r into T extends to a continuous map of the $(r+1)$-ball B^{r+1} with boundary S^r into T. Thus 0-connected means arcwise connected, 1-connected means arcwise connected and simply connected (trivial fundamental group) etc.

Theorem 4.1. *If $\mathcal{N}(G)$ is k-connected, then $\chi(G) \geq k+3$.*

This theorem has been used to prove a conjecture of Kneser concerning the chromatic number of certain graphs. Its proof depends on the Borsuk–Ulam theorem on antipodal mappings of the sphere.

Schrijver and the present author have found the following lower bound on $\tau(G)$ of a somewhat similar character. Let G be a graph and define a simplicial complex $\mathcal{M}(G)$ whose vertices are those subsets X of $V(G)$ for which both X and $V(G)-X$ span at least one line. Let the simplices of $\mathcal{M}(G)$ be those sets of such subsets which are totally ordered with respect to inclusion.

Theorem 4.2. *If $\mathcal{M}(G)$ is k-connected, then $\tau(G) \geq k+3$.*

This result generalizes to hypergraphs without any essential change. Its proof depends on the Borsuk–Ulam theorem again.

The algorithmic aspects of these topological results are very much unexplored. It is likely that the k-connectivity of $\mathcal{N}(G)$ is an NP-property for every fixed k, since it means that the k-skeleton is contractible to a single point within the $(k+1)$-skeleton, and probably this contraction can be described in polynomial time. I could not, however, work out a rigorous proof. The situation is even more complicated with $\mathcal{M}(G)$, since this has exponentially many vertices. But we may replace $\mathcal{M}(G)$ by any simplicial complex which is homotopically equivalent. Is there such a complex which has only $|V(G)|^{\text{const}}$ vertices ? Is there one which can be constructed from G in polynomial time ? Probably these questions may be answered in the affirmative using some methods like the (homotopical) Crosscut Theorem of Mather [13] or other related results on topological spaces associated with posets, lattices, etc. (we also refer to [15] and [1]).

Conclusion. We have surveyed some methods to obtain upper bounds on $\alpha(G)$ (or, equivalently, lower bounds on $\tau(G)$), which use non-trivial tools from other parts of mathematics. We tried to show that complexity considerations in connection with these methods raise some interesting questions in other fields of mathematics.

Our selection has clearly not been representative for all approaches. We have to call the reader's attention to the work of, among others, Hammer and Simeone [7], Hansen [8] and Haemers [5], but we cannot discuss their approach in detail.

References

[1] K. Baclavski and A. Björner, Fixed points in partially ordered sets, Adv. in Math. 31 (1979) 263–287.
[2] A.E. Brouwer and A. Schrijver, The blocking number of an affine plane, J. Combin. Theory Ser. A 24 (1978) 251–253.
[3] I. Csonka, L. Lovász and G. Turán, to be published.
[4] M. Grötschel, L. Lovász and A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, Combinatorica 1 (1981) 169–197.
[5] W. Haemers, On some problems of Lovász concerning the Shannon capacity of a graph, IEEE Trans. Inf. Theory 25 (1980) 231–232.
[6] G. Hajós, Über eine Konstruktion nicht n-färbbarer Graphen, Wiss. Zeitschr. Martin-Luther Univ. Halle-Wittenberg A10 (1961) 116–117.
[7] P.L. Hammer and B. Simeone, Best linear upper bounds on quadratic 0-optimization, preprint.
[8] P. Hansen, Upper bounds for the stability number of a graph, Revue Roumaine de Math. Pures Appl. 24 (1979) 1195–1199.
[9] W.W. Li and S.R. Li, Independence numbers of graphs and generators of ideals, Combinatorica 1 (1981).
[10] L. Lovász, Flats in geometries and geometric graphs, in: P. Cameron, ed., Combinatorial Surveys (Academic Press, New York, 1977) pp. 45–86.
[11] L. Lovász, Some finite basis theorems in graphs theory, in: A. Hajnal and V.T. Sós, eds., Combinatorics, Coll. Math. Soc. J. Bolyai 18 (North-Holland, Amsterdam, 1978) pp. 717–729.
[12] L. Lovász, Kneser's conjecture, chromatic number, and homotopy, J. Combin. Theory A25 (1978) 319–324.

[13] J. Mather, Invariance of the homology of a lattice, Proc. Amer. Math. Soc. 17 (1966) 1120–1124.
[14] G. Minty, On maximal independent sets of vertices in claw-free graphs, J. Combin. Theory Ser. B 28 (1980) 284–304.
[15] G.-C. Rota, On the foundations of combinatorial theory I; theory of Möbius functions, Z. Wahrsch. Verw. Geb. 2 (1964) 340–368.
[16] N. Sbihi, Algorithme de recherche d'un stable de cardinalité maximum dans un graphe sans étoile, Discrete Math. 29 (1980) 53–76.
[17] N.Z. Shor, Convergence rate of the gradient descent method with dilatation of the space, Kibernetika 2 (1970) 80–85; (English translation: Cybernetics 6 (1970) 102–108).
[18] D.B. Yudin and A.S. Nemirovskii, Informational complexity and effective methods of solution for convex extremal problems, Ekonomika i Mat. Metody 12 (1979) 357–369; (English translation: Matekon: Transl. of Russian and East European Math. Economics 13(3) 24–45).

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 225–239

SCHEDULING PROBLEMS WITH A SINGULAR SOLUTION*

Rolf H. MÖHRING

Lehrstuhl für Mathematik, Fachrichtung Operations Research, Technical University of Aachen, D-5100 Aachen, W. Germany

We consider the tasks (A) of minimizing a regular performance measure subject to resource constraints, and (B) of minimizing costs for resource requirements subject to a fixed completion time for arbitrary project networks with resource requirements. For these tasks we investigate those problems in which the optimum value is determined by the same partial order on the set of activities for all possible variations of the objective function (problems with a singular solution). It is shown that these problems constitute a large class of scheduling problems which can be recognized and solved in polynomial time. Furthermore, we give a complete characterization of all problems having a singular solution which is based on comparability graph recognition and forbidden substructures.

1. Introduction

We consider the tasks (A) of minimizing regular measures of performance subject to resource constraints, and (B) of minimizing costs for resource requirements subject to a fixed completion time in project networks with resource requirements.

Such a network is given by a partial order Θ_0 (representing the technological precedence constraints) on a set $A = \{\alpha_1, \ldots, \alpha_n\}$ of activities, a vector of activity durations $x = (x_1, \ldots, x_n)$, resource requirements c^i, $i \in I := \{1, \ldots, m\}$ ($c^i(\alpha)$ denotes the amount of resource $i \in I$ required by α), and a regular measure of performance κ (i.e., κ is a non-decreasing real-valued function of the completion times of all activities). In (A) we assume that the available amount of resource $i \in I$ is limited by a constant C_i, whereas in (B) the joint purchase of $(C_i)_{i\in I}$ units causes costs $k((C_i)_{i\in I})$ and κ must respect a given limit t_0.

Following the approach in [16, 18, 20], the solutions to problems of type (A) or (B) may be represented by partial orders Θ extending Θ_0 and fulfilling certain feasibility conditions (cf. Section 2).

This paper deals with problems with a *singular solution*, i.e., problems for

* This paper was supported by the Minister für Wissenschaft und Forschung des Landes Nordrhein-Westfalen.

which the optimum is already determined by a *single* feasible partial order Θ for all choices of κ (from a fixed class) and x in (A) and k, $c^1, \ldots, c^m$ in (B).

Problems with a singular solution are shown to constitute a class of polynomially solvable problems which contains problems of arbitrary size and whose cardinality—at least for task (A)—is larger than that of some of the common classes of job-scheduling problems (cf. Section 3).

Because of these properties, problems with a singular solution have practical importance as test examples for (heuristic) algorithms for arbitrary scheduling problems in the sense that they form a class of easily constructable and solvable problems on which a 'good' algorithm should not perform too badly.

The main interest in these problems is, however, theoretical and concerns their algorithmic and structural characterization, which reveals strong connections with comparability graph recognition and properties based on forbidden substructures. The results also demonstrate considerable symmetry lags between the at first sight rather closely related tasks (A) and (B) (cf. also [16]).

The results for (A) are presented in Section 3. Some of them, viz. that, for $\kappa = \max$, problems with a singular solution can be recognized and solved in polynomial time and that each extension of Θ_0 can occur as the singular solution of an appropriately formulated problem, were obtained by Radermacher in a more general context which will be published separately [20]. We shall give proofs for these results only as far as our approach differs from that in [20] and as far as proof methods are essential for the results on (B).

Sections 4 and 5 deal with singular solutions for (B). In Section 4, the measure of performance κ is arbitrary. Again—and this is the most important feature in common with (A)—problems with a singular solution can be recognized and solved in polynomial time. Contrary to (A), however, only those partial orders which are induced by a schedule may occur as a singular solution over Θ_0.

For the case that $\kappa = \max$ (project duration), which is treated in Section 5, only sufficient conditions for Θ to be a singular solution over Θ_0 can be given if Θ_0 is arbitrary.

If Θ_0 is *degenerate* (i.e., $\alpha \leqslant_{\Theta_0} \beta \Rightarrow \alpha = \beta$), we obtain a complete characterization of all singular solutions over Θ_0 in terms of forbidden suborders. As a consequence, only special series-parallel partial orders occur as a singular solution.

2. Basic concepts

A *project network with resource requirement* consists of

- a set $A = \{\alpha_1, \ldots, \alpha_n\}$ of the project's *activities*. They form the smallest units of the project which have to be carried out without interruption.

- a vector of *activity durations* $x=(x_1,\ldots,x_n)\in\mathbb{R}^n_>$, where $x_j=x(\alpha_j)>0$ denotes the duration of activity α_j $(j=1,\ldots,n)$.
- a partial order Θ_0 on A which describes the *technological precedence constraints* between the activities (i.e., $\alpha\leqslant_{\Theta_0}\beta$ means that α must be finished before β can be started).
- finitely many resources $i\in I:=\{1,\ldots,m\}$ and associated vectors of *resource requirements* $c^1,\ldots,c^m\in(\mathbb{N}\cup\{0\})^m$, where $c^i_j=c^i(\alpha_j)$ denotes the amount of resource $i\in I$ required by activity α_j. This amount is assumed to be constant for the full duration of α_j.

Since activities have to be carried out without interruption, each possible performance of the project is completely described by a *schedule* $T=(T(\alpha_1),\ldots,T(\alpha_n))\geqslant 0$ which gives the starting times of all activities.

Performance control is done by so-called *regular measures of performance* (cf. [5, 22]), i.e., non-decreasing functions $\kappa:\mathbb{R}^n_\geqslant\to\mathbb{R}^1$ which assign to each schedule T *performance costs* $\kappa(T)=\kappa(t_1,\ldots,t_n)\in\mathbb{R}^1$ resulting from the completion times $t_j=T(\alpha_j)+x(\alpha_j)$ $(j=1,\ldots,n)$ of the activities α_j with regard to T.

The most important measures of performance in the paper will be of the form $\kappa(t_1,\ldots,t_n)=\max\{t_1,\ldots,t_n\}$, i.e., *maximum completion time* or *project duration*, or the form $\kappa=f\circ\max$, where $f:\mathbb{R}^1\to\mathbb{R}^1$ is non-decreasing.

For each schedule T, let $C_i(T)$ denote the maximum amount of resource i required while the project is performed according to T, i.e. $C_i(T):=\max\{\sum_{T(\alpha)\leqslant t<T(\alpha)+x(\alpha)} c^i(\alpha)\mid t\geqslant 0\}$.

We then consider two problems, cf. [16, 20].

Problem A (the problem of scarce resources). The available amount of each resource $i\in I$ is limited by a constant $C_i\in\mathbb{N}$ with $c^i(\alpha)\leqslant C_i$ for all $\alpha\in A$. Minimize $\kappa(T)$ subject to all schedules T for which $C_i(T)\leqslant C_i$ $(i=1,\ldots,m)$.

Problem B (the problem of scarce time). Resources are available time-independently at costs $k(C_1,\ldots,C_m)$ for a joint purchase $(C_1,\ldots,C_m)$ of C_i units of resource i, $i\in I$. The measure of performance may not, however, exceed a given limit t_0. Minimize the costs $k(C_1(T),\ldots,C_m(T))$ of resource requirements subject to all schedules T for which $\kappa(T)\leqslant t_0$.

Instead of the above standard representation of these problems by means of feasible schedules we shall use a representation based on 'feasible' partial orders on A. This approach, which has also been applied in the disjunctive graph concept in job shop scheduling [1, 8, 9], allows a clear distinction to be made between feasibility domain and objective function.

The basic idea is that a schedule T (and, more generally, any function $T:A\to\mathbb{R}^1_\geqslant$) together with a duration vector x induces a partial order $\Theta(T,x)$ on A by putting (observe that $x(\alpha)>0$ for all α)

$$\alpha <_{\Theta(T,x)} \beta :\Leftrightarrow T(\alpha) + x(\alpha) \leqslant T(\beta).$$

$\Theta(T, x)$ is said to be *induced* by T and x.

If T is a schedule for Θ and x (i.e., $\alpha <_\Theta \beta \Rightarrow T(\alpha) + x(\alpha) \leqslant T(\beta)$), then $\Theta(T, x)$ is obviously an *extension* of Θ, i.e., $\Theta(T, x) \supseteq \Theta$ (meaning that $\alpha \leqslant_\Theta \beta$ implies $\alpha \leqslant_{\Theta(T,x)} \beta$).

Given a partial order Θ on A, lèt $\mathfrak{u}(\Theta)$ be the set of all anti-chains (independent sets) of Θ. Then $\phi_\Theta(w) := \max\{\Sigma_{\alpha \in U} w(\alpha) \mid U \in \mathfrak{u}(\Theta)\}$ denotes the maximum weight of an anti-chain of Θ with regard to a weight vector $w = (w(\alpha_1), \ldots, w(\alpha_n))$.

Proposition 2.1 [1, 18]. *Let T be a schedule for Θ_0 and x. Then $C_i(T) = \phi_{\Theta(T,x)}(c^i)$ $(i = 1, \ldots, m)$, i.e., the amount of resource i required if the project is scheduled with regard to T equals the maximum weight of an anti-chain of the induced partial order with regard to the weight vector c^i.*

In contrast to graphs, where the corresponding problem is NP-complete, we can here compute $\phi_\Theta(w)$ in polynomial bounded time by flow methods.

Proposition 2.2 [7, 10]. *Let Θ be a partial order on A. Let $G = (V, E)$ be the directed graph with vertex set $V = A \cup \{s, t\}$ $(s, t \notin A)$ and edge set $E = \{(\alpha, \beta) \mid \alpha <_\Theta \beta\} \cup \{s\} \times A \cup A \times \{t\}$. Then $\phi_\Theta(w)$ equals the minimum flow value in G when the nodes have lower capacities $w(\alpha)$ and edge capacities are infinite.*

As a consequence, $\phi_\Theta(w)$ can be computed in $O(n^3)$ time by standard flow methods [13]. (A representation of $\phi(w)$ as a maximum flow value is given in [8]. It is, however, not polynomial in the input data.)

Given Θ and x, let $\mathrm{ES}_\Theta[x]$ denote the *earliest start schedule*, i.e., the (componentwise) least schedule for Θ and x. It may be computed iteratively in $O(n^2)$ time according to

$$\mathrm{ES}_\Theta[x](\alpha) = \begin{cases} 0, & \alpha \text{ minimal in } \Theta, \\ \max\limits_{\beta <_\Theta \alpha} [\mathrm{ES}_\Theta[x](\beta) + x(\beta)], & \text{otherwise.} \end{cases}$$

Thus, $\mathrm{ES}_\Theta[x] \leqslant T$ for all schedules T for Θ, x. Since performance measures are non-decreasing, we also obtain $\kappa(\Theta, x) := \kappa(\mathrm{ES}_\Theta[x]) \leqslant \kappa(T)$ for all schedules T for Θ and x.

This observation, together with Proposition 2.1, gives the desired representation of Problems A and B, cf. [16, 18, 20]:

(A) $\min\{\kappa(\Theta, x) \mid \Theta \in \mathcal{O}_A\}$ with $\mathcal{O}_A = \{\Theta \mid \Theta_0 \subseteq \phi_\Theta(c^i) \leq C_i,\ i = 1, \ldots, m\}$,
(B) $\min\{k(\phi_\Theta(c^1), \ldots, \phi_\Theta(c^m)) \mid \Theta \in \mathcal{O}_B\}$ with $\mathcal{O}_B = \{\Theta \mid \Theta_0 \subseteq \Theta,\ \kappa(\Theta, x) \leq t_0$.

The partial orders from $\mathcal{O}_A$ or $\mathcal{O}_B$ are called *feasible* for (A) or (B).

The feasibility domains $\mathcal{O}_A$ and $\mathcal{O}_B$ are finite and convex (i.e., if $\Theta_1 \subseteq \Theta \subseteq \Theta_2$ and if Θ_1, Θ_2 are feasible, then Θ is feasible as well) subsets of the semi-lattice $\mathcal{O}(\Theta_0)$ of all extensions of Θ_0 ordered with regard to inclusion. Since the objective functions $\Theta \to \kappa(\Theta, x)$ for (A) and $\Theta \to k(\phi_\Theta(c^1), \ldots, \phi_\Theta(c^m))$ for (B) are non-decreasing and non-increasing on $\mathcal{O}(\Theta_0)$, respectively, the optimum is already attained by the minimal feasible partial orders in (A) and the maximal feasible partial orders in (B).

Of special interest are minimal or least *dominating* sets of feasible partial orders, i.e., minimal or least subsets $\mathcal{O}$ of $\mathcal{O}_A$ or $\mathcal{O}_B$ which already determine the optimum for a fixed class of performance measures κ and all possible choices of x in (A) and for all possible choices of k and $c^1, \ldots, c^m$ in (B).

For general results on dominating sets (and also for algorithmic methods based on this approach) cf. [18, 20] for (A) and [16] for (B).

In this paper we will concentrate on problems with a *singular solution*, i.e., problems for which the dominating sets are singletons. For these problems, the optimum is already determined by a *single* feasible partial order Θ for all choices of κ (from a fixed class) and x in (A) and k, $c^1, \ldots, c^m$ in (B).

3. Singular solutions for (A)

Let Θ_0 and $(c^i, C_i)_{i \in I}$ be fixed. An anti-chain $U \in \mathrm{u}(\Theta_0)$ is called *adaptive* if there exists a feasible partial order Θ with $U \in \mathrm{u}(\Theta)$. Now to each $U \in \mathrm{u}(\Theta_0)$ with $|U| \geq 2$ there obviously exists an extension Θ_u of Θ_0 such that U is the only non-singleton anti-chain of Θ_u. Thus $U \in \mathrm{u}(\Theta_0)$ is adaptive iff $\Sigma_{\alpha \in U}\, c^i(\alpha) \leq C_i$ for all $i \in I$.[1]

Let $\mathcal{N} = \{N_1, \ldots, N_r\}$ denote the system of $\subseteq$-minimal non-adaptive anti-chains of Θ_0. It then follows easily that a partial order $\Theta \supseteq \Theta_0$ is feasible iff Θ introduces a precedence constraint $\alpha_j \leq_\Theta \beta_j$ on each set $N_j \in \mathcal{N}$.

Call two problems given by Θ_0, $(c^i, C_i)_{i \in I}$ and Θ_0', $((c^i)', C_i')_{i \in I'}$ *not essentially distinct* w.r.t. a class $\mathcal{K}$ of performance measures if the optima coincide for all possible choices of $\kappa \in \mathcal{K}$ and x.

It is shown in [20] that if $\mathcal{K}$ contains $\kappa = \max$ and $\Theta_0 = \Theta_0'$, then the two problems are essentially distinct iff they have distinct systems $\mathcal{N}$.

So the problems with a singular solution over a fixed technological partial

[1] Note the difference with the analogous notion for (B) in Section 4.

order Θ_0 may be classified according to their system $\mathcal{N}$. It turns out that only very special systems $\mathcal{N}$ occur, which correspond to certain comparability graphs (cf. [7]) containing the comparability graph $G(\Theta_0)$ of Θ_0 as a subgraph.

Theorem 3.1. *Let the problem given by Θ_0, $(c^i, C_i)_{i\in I}$ have a singular solution for a class of performance measures containing $\kappa = \max$. Then there exists a uniquely determined comparability graph G such that*

(i) *G has a transitive orientation extending Θ_0;*

(ii) *$\mathcal{N}$ consists of all edges $\{\alpha, \beta\}$ of G which do not belong to $G(\Theta_0)$.*

Proof. Let Θ be a singular solution of the problem. Then $\mathfrak{u}(\Theta) \supseteq \mathfrak{u}(\Theta')$ for each feasible partial order Θ'. Otherwise there would exist an anti-chain $\{\beta_1, \beta_2\} \in \mathfrak{u}(\Theta')\setminus \mathfrak{u}(\Theta)$, which means that Θ would not be an optimal solution for $\kappa = \max$ and x defined by

$$x(\alpha) = \begin{cases} 1, & \alpha = \beta_1, \beta_2 \\ \varepsilon, & \text{otherwise} \end{cases} \quad \text{with } \varepsilon < \frac{1}{2|A|},$$

since $\kappa(\Theta', x) < 2 \leq \kappa(\Theta, x)$.

Thus each $\{\alpha, \beta\} \in \mathfrak{u}(\Theta_0)\setminus \mathfrak{u}(\Theta)$ is non-adaptive and a member of $\mathcal{N}$. As Θ is feasible, each $N_j \in \mathcal{N}$ contains a pair $\{\alpha, \beta\} \in \mathfrak{u}(\Theta_0)\setminus \mathfrak{u}(\Theta)$. Since the N_j are minimal non-adaptive, we obtain that $\mathcal{N} = \{\{\alpha, \beta\} \in \mathfrak{u}(\Theta_0) \mid \alpha <_\Theta \beta\}$.

The comparability graph G of Θ obviously fulfils (i) and (ii). It is also unique, since any other comparability graph G' fulfilling (i) and (ii) is the comparability graph of a feasible poset Θ'. Then $\mathfrak{u}(\Theta') \subseteq \mathfrak{u}(\Theta)$ and (ii) yield $\mathfrak{u}(\Theta') = \mathfrak{u}(\Theta)$, i.e., $G = G'$. $\square$

Thus each singular solution Θ of the problem must be a transitive orientation of G which extends Θ_0. In general, there are several such orientations, say $\Theta_1, \ldots, \Theta_r$. The question whether one (or more specific, which) of them is a singular solution depends on the class of performance measures considered.

For instance, if we take performance measures of the form

$$\kappa(t_1, \ldots, t_n) = \begin{cases} 0 & (t_1, \ldots, t_n) \leq (d_1, \ldots, d_n) \\ 1 & \text{otherwise} \end{cases} \quad (\text{'tardiness costs'}),$$

no singular solution will exist, since by varying κ appropriately we can obtain for each set $\{\alpha, \beta\} \in \mathcal{N}$ that $\alpha \leq \beta$ is preferred to $\beta \leq \alpha$ and vice-versa.

The opposite case, when each Θ_j, $j = 1, \ldots, r$ is a singular solution, occurs if the performance measures are *invariant* with regard to all partial orders having the same comparability graph (i.e., $G(\Theta_1) = G(\Theta_2) \Rightarrow \kappa(\Theta_1, x) = \kappa(\Theta_2, x)$ for all x).

The most important invariant performance measure is $\kappa = \max$. In this case we have the well-known equality (cf. [11])

$$\kappa(\Theta, x) = \max\left\{ \sum_{\alpha \in K} x(\alpha) \mid K \in \mathfrak{c}(\Theta) \right\} =: \Lambda_\Theta(x)$$

where $\mathfrak{c}(\Theta)$ denotes the system of totally ordered subsets (chains) of Θ. $\Lambda_\Theta(x)$ can be computed in $O(n^2)$ time using standard critical path methods.

Furthermore, $\mathfrak{c}(\Theta_1) = \mathfrak{c}(\Theta_2)$ iff $G(\Theta_1) = G(\Theta_2)$, which proves that $\kappa = \max$ and performance measures of the form $\kappa = f \circ \max$ are invariant. Lemma 3.2 shows that these are the only invariant performance measures.

Lemma 3.2. *Let κ be invariant w.r.t. partial orders having the same comparability graph. Then $\kappa = f \circ \max$, where $f: \mathbb{R}^1 \to \mathbb{R}^1$ is non-decreasing.*

Proof. All linear orders on A have the same comparability graph. Put $x = (d, 0, \ldots, 0)$ (zero durations are allowed here for simplicity's sake). Then the earliest completion time vectors $t = (t_1, \ldots, t_n)$ of the linear orders on A w.r.t. x are given by all non-zero extreme points of the cube $[0, d]^n$. Thus κ is constant on all faces of $[0, d]^n$ not containing 0. This means that $\kappa(t) = f(\max\{t_1, \ldots, t_n\}) = f(d)$. □

Theorem 3.3. *For $\kappa = f \circ \max$, each extension Θ of Θ_0 can be the singular solution of a problem over Θ_0. The number of essentially distinct problems over Θ_0 with a singular solution equals the number of comparability graphs G which have a transitive orientation extending Θ_0.*

Proof. Theorem 3.1 and the fact [18] that each system $\{N_1, \ldots, N_s\}$ of non-singleton anti-chains of Θ_0 is the system $\mathcal{N}$ for appropriately defined $(c^i, C_i)_{i \in I}$. □

Between these two extremes, viz. tardiness costs and invariant performance measures, there are performance measures which distinguish between all possible orientations $\Theta_1, \ldots, \Theta_r$ of G. To them belong all performance measures of the form $\kappa = \kappa_1 + \kappa_2$, where κ_1 is invariant and $\kappa_2(t_1, \ldots, t_n) = f(t_j)$, $j \in \{1, \ldots, n\}$, $f: \mathbb{R}^1 \to \mathbb{R}^1$ non-decreasing.

As a consequence of Theorem 3.1, problems with a singular solution can for the case $\kappa = \max$ be recognized and solved in $O(|I| \cdot n^3)$ time by the following algorithm, which makes use of the $O(n^3)$-algorithm of Golumbic [7] for comparability graph recognition and orientation.

Algorithm 3.4. Let Θ_0, $(c^i, C_i)_{i \in I}$ and $\kappa = \max$ be given.

Step 1. Compute all pairs $\{\alpha, \beta\} \in (\Theta_0)$ which are non-adaptive, i.e., for which $c^i(\alpha) + c^i(\beta) > C_i$ for some $i \in I$.

Step 2. Add all non-adaptive pairs $\{\alpha, \beta\}$ as edges to the comparability graph $G(\Theta_0)$ of Θ_0. Try to extend the given transitive orientation Θ_0 of $G(\Theta_0)$ to a transitive orientation of the new graph by Golumbic's algorithm [7]. If no such orientation exists, the problem does (because of Theorem 3.1) not have a singular solution. Terminate. Otherwise go to Step 3.

Step 3. Test whether the orientation Θ found in Step 2 is feasible by testing $\phi_\Theta(c^i) \leqslant C_i$, $i = 1, \ldots, m$. This can be done by the flow methods described in Proposition 2.2 in $O(|I| \cdot n^3)$ time. If Θ is feasible, then it is a singular solution of the problem because of Theorem 3.1. Go to Step 4. Otherwise the problem has no singular solution. Terminate.

Step 4. Compute $\kappa(\Theta, x) = \Lambda_\Theta(x)$, which is the optimum value.

One would, of course, like to know how large this class of polynomially solvable problems really is, in particular in comparison with other well-known classes of polynomially solvable scheduling problems. The basis for such a comparison is given by the notion of *essentially distinct problems* introduced above, and Theorem 3.3. We will make such a comparison for the case that Θ_0 is degenerate (i.e., problems without precedence constraints) with the classes defined by the conditions (Q_1) and (Q_2) below, which contain most of the well-known polynomially solvable, non-preemptive machine-scheduling problems (cf. [9, 14]):

(Q_1) $\begin{cases} I = \{1\}, \quad c_1(\alpha) \in \{0, 1\} \text{ for all } \alpha \in A \\ \text{(only one type of resource, each activity requires at most one} \\ \text{unit; contains the } \textit{parallel machine problems} \text{ in the sense of} \\ \text{[9, 14]);} \end{cases}$

(Q_2) $\begin{cases} \text{I arbitrary}, \quad c_i(\alpha) \in \{0, 1\} \text{ for all } \alpha \in A \text{ and } i \in I, \\ c_i(\alpha) = 1 \Rightarrow c_j(\alpha) = 0 \text{ for all } j \in I,\ j \neq i \text{ and all } \alpha \\ \text{(several resource types, each activity requires only one resource} \\ \text{type and at most one unit of it; contains the } \textit{unrelated machine} \\ \textit{problems} \text{ in the sense of [9, 14]).} \end{cases}$

The exact numbers $Q_1(n)$ and $Q_2(n)$ of all essentially distinct problems of the respective classes over the degenerate partial order Θ_0 on n activities are given in [18, 20].

$$Q_1(n) = (n-2)2^{n-1} + 2\,,$$

$$Q_2(n) = \sum_{\substack{\delta_l \in \mathbb{N} \cup \{0\},\ l = 1, \ldots, n \\ 1 \cdot \delta_1 + 2 \cdot \delta_2 + \cdots + n \cdot \delta_n = n}} \frac{n!}{\prod_{l=1}^{n} (l!)^{\delta_l} \cdot \delta_l!} \prod_{l=1}^{n} \max[1, (l-1)^{\delta_l}]\,.$$

Table 1
The numbers $Q_1(n)$, $Q_2(n)$, $S(n)$, $P(n)$ for $n = 1, \ldots, 10$
(numbers given in exponential representation denote lower bounds)

n	1	2	3	4	5	6	7	8	9	10
$Q_1(n)$	1	2	3	18	50	130	322	770	1794	4098
$Q_2(n)$	1	2	6	21	85	395	2051	11 690	72 458	484 559
$S(n)$	1	2	8	64	1012	30 164	1 527 346	119 369 968	2.38×10^8	4.76×10^8
$P(n)$	1	2	9	114	6894	7 785 062	3.44×10^{10}	1.18×10^{21}	8.51×10^{37}	7.24×10^{75}

From [17] and Theorem 3.3 we obtain the following bounds for the number $S(n)$ of essentially distinct problems with a singular solution over the degenerate partial order Θ_0 on n activities:

$$2^{n^2/4+3n/2-3\log_2 n-1} \leq S(n) \leq 2^{n^2/4+n\log_2 n-1}\,.$$

Clearly, $\lim_{n\to\infty} Q_1(n)/S(n) = \lim Q_2(n)/S(n) = 0$, i.e., within the class of all essentially distinct scheduling problems, conditions Q_1 and Q_2 turn out to be much more restrictive than that of having a singular solution, although, of course, with regard to the number $P(n)$ of *all* essentially distinct problems, also $\lim S(n)/P(n) = 0$ (cf. also Table 1).

4. Singular solutions for (B), general case

In order to accentuate the difference with (A), we shall proceed parallel to Section 2 as far as possible. Of course, all notions such as feasible, adaptive, etc. are now to be interpreted in the context of (B).

Let Θ_0, κ, x and t_0 be fixed. A subset B of A is called *adaptive* if there exists a feasible partial order Θ such that $B \in \mathfrak{c}(\Theta)$. Let $\mathcal{N} = \{N_1, \ldots, N_r\}$ denote the system of $\subseteq$-minimal non-adaptive subsets from A. Then, for each feasible partial order $\Theta \supseteq \Theta_0$, $\supseteq(\Theta)$ contains a pair $\{\alpha_j, \beta_j\}$ from each $N_j \in \mathcal{N}$, i.e., $N_j \notin \mathfrak{c}(\Theta)$ for all $j = 1, \ldots, r$. (The converse holds if $\kappa = \max$.)

Different to (A), however, there is no way to construct $\mathcal{N}$ directly (i.e., without knowing all feasible partial orders) from Θ_0, κ, x and t_0, even in the case $\kappa = \max$ (cf. [16]).

Theorem 4.1. *Let the problem given by Θ_0, κ, x and t_0 have a singular solution. Then there is a uniquely determined comparability graph G such that*

(i) *G has a transitive orientation extending Θ_0;*

(ii) *$\mathcal{N}$ consists of all pairs $\{\alpha, \beta\}$ which are not an edge of G.*

Proof. Let Θ be a singular solution of the problem. Then $\mathfrak{c}(\Theta) \supseteq \mathfrak{c}(\Theta')$ for each feasible partial order Θ'. Otherwise there would exist a chain $\{\beta_1, \beta_2\} \in$

$\mathfrak{c}(\Theta')\setminus \mathfrak{c}(\Theta)$, which means that Θ would not be an optimal solution for

$$I = \{1\}, \qquad c^1(\alpha) = \begin{cases} 1, & \alpha = \beta_1, \beta_2, \\ 0, & \text{otherwise}, \end{cases} \quad \text{and } k(C_1) = C_1,$$

since $k(\phi_{\Theta'}(c^1)) = 1$, $k(\phi_\Theta(c^1)) = 2$.

The rest of the proof is similar to that of Theorem 3.1, with reversed roles of $\mathfrak{c}(\Theta)$ and $\mathfrak{u}(\Theta)$.[2] □

Since $G(\Theta_1) = G(\Theta_2)$ iff $\mathfrak{u}(\Theta_1) = \mathfrak{u}(\Theta_2)$, the objective function $\Theta \to k(\phi_\Theta(c^1), \ldots, \phi_\Theta(c^m))$ is invariant with regard to posets $\Theta_1, \ldots, \Theta_r$ having the same comparability graph, so that (contrary to (A)) no restriction with regard to the objective function has to be made in order to make one of the Θ_j a singular solution.

The main difference to (A) is that not all partial orders $\Theta \supseteq \Theta_0$ occur as a singular solution, but only those which are induced by a schedule (cf. Section 2).

This is due to the fact that, if Θ is feasible for Θ_0, κ, x and t_0 (i.e., $\Theta_0 \subseteq \Theta$ and $\kappa(\Theta, x) \le t_0$), then $\Theta^* := \Theta(\mathrm{ES}_\Theta[x], x)$ is also feasible and $\Theta \subseteq \Theta^*$. Now if $\Theta' \subsetneq \Theta^*$, it is easy to construct k, $c^1, \ldots, c^m$ such that $k(\phi_{\Theta^*}(c^1), \ldots, \phi_{\Theta^*}(c^m)) < k(\phi_{\Theta'}(c^1, \ldots, \phi_{\Theta'}(c^m))$. So we obtain the following lemma.

Lemma 4.2. *Let $\Theta \supseteq \Theta_0$ be a singular solution for Θ_0, κ, x and t_0. Then $\Theta = \Theta(\mathrm{ES}_\Theta[x], x)$ and $\mathcal{N} = \{\{\alpha, \beta\} \mid \alpha \neq \beta, \{\alpha, \beta\} \in \mathfrak{u}(\Theta)\}$.*

The following lemma shows that any schedule-induced partial order can already be induced in such a way that certain additional properties hold.

Lemma 4.3. *Let Θ be induced by T and x. Then there exists a duration vector z such that*

(i) *Θ is induced by $\mathrm{ES}_\Theta[z]$ and z,*

(ii) *each activity α is critical with regard to Θ and z, i.e., its prolongation enlarges the project duration $\Lambda_\Theta(z)$.*

Proof. z may be constructed as follows: Put

$$y := \begin{cases} T(\alpha) + x(\alpha), & \alpha \text{ minimal in } \Theta, \\ T(\alpha) + x(\alpha) - \max\limits_{\beta <_\Theta \alpha} [T(\beta) + x(\beta)], & \text{otherwise}, \end{cases}$$

[2] For the case $\kappa = \max$, there are many symmetries between (A) and (B), which lead to duality relations between them by means of conjugation of partial orders in the sense of [6] (cf. [16]).

$$z := \begin{cases} \Lambda_\Theta(x) - \mathrm{ES}_\Theta[y](\alpha), & \alpha \text{ maximal in } \Theta, \\ \min\limits_{\alpha <_\Theta \beta} \mathrm{ES}_\Theta[y](\beta) - \mathrm{ES}_\Theta[y](\alpha), & \text{otherwise}. \end{cases}$$

Obviously y is defined in such a way that, for all α, $x(\alpha) \leq y(\alpha)$, $\mathrm{ES}_\Theta[y](\alpha) \leq T(\alpha)$ and $\mathrm{ES}_\Theta[y](\alpha) + y(\alpha) = T(\alpha) + x(\alpha)$. Then $\Theta \subseteq \Theta(\mathrm{ES}_\Theta[y], y) =: \Theta^*$. If $\alpha <_{\Theta^*} \beta$, then $T(\alpha) + x(\alpha) = \mathrm{ES}_\Theta[y](\alpha) + y(\alpha) \leq \mathrm{ES}_\Theta[y](\beta) \leq T(\beta)$, i.e., $\alpha <_\Theta \beta$. Thus $\Theta = \Theta^*$. In constructing z from y, each activity duration is prolonged so as to satisfy the equation $\mathrm{ES}_\Theta[y](\alpha) + z(\alpha) = \min_{\alpha <_\Theta \beta} \mathrm{ES}_\Theta[y](\beta)$. Thus, the rest of Lemma 4.3 is obvious. □

Theorem 4.4. *Let $\Theta \supseteq \Theta_0$ be schedule-induced. Then there exist κ, x, t_0 such that Θ is a singular solution for Θ_0, x and t_0.*

Proof. Let x be the duration vector z from Lemma 4.3, $t_0 := 1$ and

$$\kappa(t_1, \ldots, t_n) := \begin{cases} 0, & t_j \leq \mathrm{ES}_\Theta[x](\alpha_j) + x(\alpha_j) \text{ for all } j, \\ 1, & \text{otherwise}. \end{cases}$$

Then obviously only the sets from $\mathrm{c}(\Theta)$ are adaptive, i.e. $\mathcal{N}$ fulfils the conditions in Theorem 3.1. □

It should be noted that only an asymptotically vanishing proportion of partial orders on n elements is schedule-induced. This follows from the fact that schedule-induced partial orders can be characterized by forbidden suborders [21]. Since (as in the theory of properties of almost all graphs [2]) almost all partial orders contain any specified partial order as a suborder [4], almost no partial order is schedule-induced.

This means that, asymptotically, only a vanishing fraction of partial orders may occur as a singular solution for a problem of type (B), whereas for task (A), *all* partial orders may occur.

The following algorithm permits recognition and solution of problems with a singular solution in polynomial time, provided that the computation of κ is possible in polynomial time.

Algorithm 4.5. Let Θ_0, κ, x and t_0 be given.

Step 1. Compute all pairs $\{\alpha, \beta\} \in \mathrm{u}(\Theta_0)$ which are adaptive. To this end, let $\Theta_{\alpha\beta}$ denote the extension of Θ_0 obtained by introducing the relation $\alpha \leq \beta$, i.e., by adding all pairs $\gamma \leq \delta$ with $\gamma \leq_{\Theta_0} \alpha$ and $\beta \leq_{\Theta_0} \delta$. Then $\{\alpha, \beta\}$ is adaptive iff $\kappa(\Theta_{\alpha\beta}, x) \leq t_0$ or $\kappa(\Theta_{\beta\alpha}, x) \leq t_0$.

Step 2. Add all adaptive pairs $\{\alpha, \beta\}$ as edges to the comparability graph $G(\Theta_0)$ of Θ_0. Try to extend the orientation Θ_0 of $G(\Theta_0)$ to a transitive orientation Θ of the new graph by Golumbic's algorithm. If no such orientation exists, the problem does not have a singular solution. Terminate. Otherwise go to Step 3.

Step 3. If $\kappa(\Theta, x) > t_0$, Θ is not feasible and there exists no singular solution. Terminate. Otherwise go to Step 4.

Step 4. Θ is a singular solution of the problem. The objective value $k(\phi_\Theta(c^1), \ldots, \phi_\Theta(c^m))$ is computed by the flow methods described in Proposition 2.2.

5. Singular solutions for (B), case κ = max

If κ is restricted to the case κ = max, Theorem 4.4 remains valid for the case that x is the vector z from Lemma 4.3, $t_0 = \Lambda_\Theta(x)$ and $\mathrm{ES}_{\Theta_0}[x] = \mathrm{ES}_\Theta[x]$, but not in general.

Example 5.1. Let Θ be the partial order on $A = \{1, \ldots, 6\}$ given by the arrow diagram (cf. [11]) in Fig. 1.

Then $x = (2, 5, 7, 1, 2, 5)$ is a duration vector in the sense of Lemma 4.3. Let $\kappa :=$ max and $t_0 := \Lambda_\Theta(x) = 11$. Then $\mathrm{ES}_{\Theta^*}[x] = \mathrm{ES}_\Theta[x]$ for all Θ^* such that $\Theta_1 \subseteq \Theta^* \subseteq \Theta$ (cf. Fig. 1 for Θ_1). Thus Θ is a singular solution for each such Θ^*. Θ is also a singular solution for all 5 partial orders Θ_0 which result from Θ by deleting a pair $\alpha \leqslant \beta$, where β covers α, although $\mathrm{ES}_{\Theta_0}[x] \neq \mathrm{ES}_\Theta[x]$ is possible. Θ is, however, no longer a singular solution if $1 \leqslant_{\Theta_0} 3$ and $1 \leqslant_{\Theta_0} 4$ are deleted simultaneously.

The general problem for κ = max, viz. to give a characterization of all singular solutions for arbitrary Θ_0, is still unsolved. We shall give a solution for the case that Θ_0 is degenerate (i.e., $\alpha \leqslant_\Theta \beta$ implies $\alpha = \beta$). In this case, a subset B of A is adaptive for Θ_0, x and t_0 iff $\Sigma_{\alpha \in \beta}\, x(\alpha) \leqslant t_0$.

Call a partial order Θ on A *series-parallel* if it is built up recursively as follows:

(1) The partial orders on two points are series-parallel.

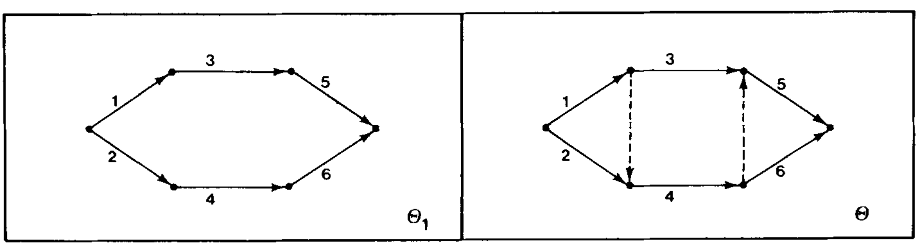


Fig. 1. The partial orders of Example 5.1.

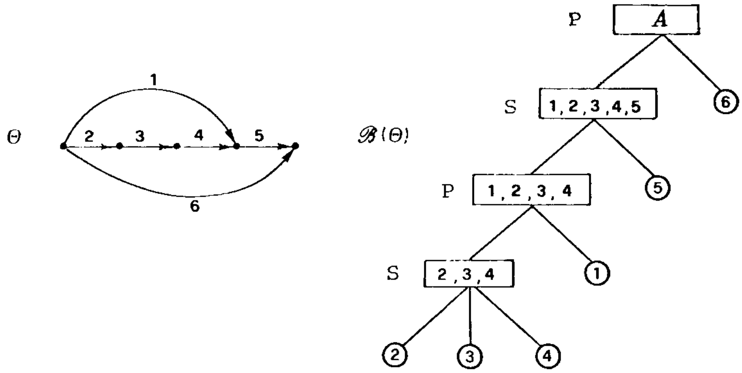


Fig. 2. A series-parallel partial order and its composition tree.

(2) If Θ_1 on A_1 and Θ_2 on A_2 with $A_1 \cap A_2 = \emptyset$ are series-parallel, then $\Theta^S = \Theta_1 \cup \Theta_2 \cup (A_1 \times A_2)$ on $A_1 \cup A_2$ and $\Theta^P = \Theta_1 \cup \Theta_2$ on $A_1 \cup A_2$, too. Θ^S and Θ^P are said to be built up from Θ_1 and Θ_2 by *series composition* and *parallel composition*, respectively. Θ_1 and Θ_2 are called *series blocks of* Θ^S and *parallel blocks* of Θ^P.

A series-parallel partial order Θ on A can be fully described by a rooted tree, the *composition tree* $\mathscr{B}(\Theta)$.

The root of $\mathscr{B}(\Theta)$ is A. Each node B of $\mathscr{B}(\Theta)$ is a subset of A. If $|B| \geq 2$, then B is assigned the label P or S depending on whether the suborder $\Theta \mid B$ of Θ induced by B is built up by parallel or series composition in the sense of (2). The blocks of the finest partition of $\Theta \mid B$ into parallel or series blocks are the successor nodes of the node B (cf. Fig. 2).

$\mathscr{B}(\Theta)$ is a special form of the composition tree for arbitrary partial orders (cf. [3]). A slightly different definition of composition tree for series-parallel partial orders has also been given in [23].

Another characterization has been obtained by forbidden suborders [11, Theorem 18.6] and [23]: Θ is series-parallel iff it does not contain a suborder isomorphic to Θ_2 in Fig. 3.

Because of this forbidden suborder, again (as for schedule-induced orders) almost no partial order will be series-parallel. This also follows from the results obtained in [15].

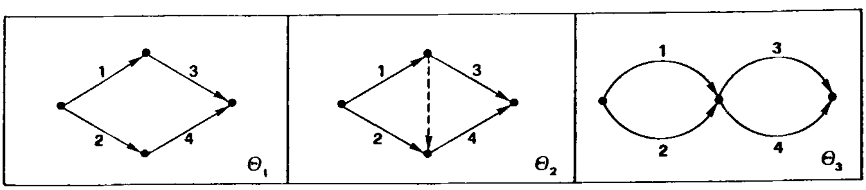


Fig. 3. The partial orders of Theorem 5.2.

Theorem 5.2. *Let Θ_0 be degenerate and $\kappa = \max$. Then the following statements are equivalent.*

(1) *Θ is a singular solution of a problem over Θ_0 for $\kappa = \max$.*

(2) *Θ does not contain any of the partial orders of Fig.* 3 *as a suborder.*

(3) *Θ is series-parallel and each node of $\mathcal{B}(\Theta)$ has at most one non-singleton successor.*

Proof. (1)$\Rightarrow$(2) Assume that Θ contains Θ_1, Θ_2, or Θ_3 from Fig. 3 as a suborder. Then, by Theorem 4.1, $\{1,2\}$ and $\{3,4\}$ belong to $\mathcal{N}$. Since Θ_0 is degenerate, $x_1+x_2>t_0$ and $x_3+x_4>t_0$. On the other hand, $\{1,3\}$ and $\{2,4\}$ are adaptive, which yields $x_1+x_3 \leqslant t_0 < x_1+x_2$ and $x_2+x_4 \leqslant t_0 < x_3+x_4$, i.e., $x_3<x_2$ and $x_2<x_3$, a contradiction.

(2)$\Rightarrow$(3) Since Θ does not contain Θ_2, Θ is series-parallel because of the above characterization. Since it does not contain Θ_1 and Θ_3, a node with the label P or S can at most have one non-singleton successor.

(3)$\Rightarrow$(1) We show by induction on the depth of $\mathcal{B}(\Theta)$ that there exist x and t_0 such that $x(\alpha)+x(\beta)>t_0$ for all $\{\alpha,\beta\}\in \mathrm{u}(\Theta)$ and $\Lambda_{\Theta|B}(x)=t_0$ for all series blocks B of Θ. This is trivial for depth 1. So let one of the successors $B_1,\dots,B_m$ of A be non-singleton, say B_1. Then $|B_2|=\cdots=|B_m|=1$, and by the inductive hypothesis there exist x^* and t_0^* for $\Theta \mid B_1$ with the above properties. Then x and t_0 are given by

$$x(\alpha):=\begin{cases} x^*(\alpha) & \text{if } \alpha\in B_1,\\ t_0^* & \text{otherwise,}\end{cases} \qquad t_0:=t_0^* \text{ if } A \text{ is labeled with P}$$

and

$$x(\alpha):=\begin{cases} m\cdot x^*(\alpha) & \text{if } \alpha\in B_1\\ 1 & \text{otherwise}\end{cases} \qquad t_0:=mt_0^*+(m-1) \text{ if } A \text{ is labeled with S.}$$

Thus (1) follows from Theorem 4.1 and the fact that a set B' is adaptive iff $\Sigma_{\alpha\in B'}x(\alpha)\leqslant t_0$. $\square$

For Θ from Fig. 2, we would obtain $x=(6,2,2,2,1,7)$ and $t_0=7$.

Acknowledgement

I would like to thank the referee for his/her valuable comments and suggestions.

References

[1] E. Balas, Project scheduling with resource constraints, in: E.M.L. Beale, ed., Appl. of Math. Progr. (The English University Press, London, 1971).

[2] A. Blass and F. Harary, Properties of almost all graphs and complexes, J. Graph Theory 3 (1979) 225–240.
[3] H. Buer and R.H. Möhring, A fast algorithm for the decomposition of graphs and posets, submitted; extended abstract appeared in Methods of Oper. Res. 37 (1980) 259–264.
[4] K.J. Compton, private communication.
[5] R.W. Conway, W.L. Maxwell and L.W. Miller, Theory of Scheduling (Addison-Wesley, Reading, MA, 1967).
[6] B. Dushnik and E.W. Miller, Partially ordered sets, Amer. J. Math. 63 (1941) 600–610.
[7] M.C. Golumbic, The complexity of comparability graph recognition and coloring, Computing 18 (1977) 199–208.
[8] S. Gorenstein, An algorithm for project (job) sequencing with resource constraints, Oper. Res. 20 (1972) 835–850.
[9] R.L. Graham, E.L. Lawler, J.K. Lenstra and A.H.G. Ronnooy Kan, Optimization and approximation in deterministic sequencing and scheduling: a survey, Ann. Discrete Math. 5 (1979) 287–326.
[10] R. Kaerkes and B. Leipholz, Generalized network functions in flow networks, Methods of Oper. Res. 27 (1977) 225–273.
[11] R. Kaerkes and R.H. Möhring, Vorlesungen über Ordnungen und Netzplantheorie, Schriften zur Informatik and Angewandten Mathematik 45, Techn. Univ. of Aachen, 1978.
[12] R. Kaerkes, R.H. Möhring, W. Oberschelp, F.J. Radermacher and M.M. Richter, Netzplanoptimierung: Deterministische und stochastische Scheduling-Probleme über geordneten Strukturen, Lecture Notes in Economics and Mathematical Systems (Springer, Heidelberg, 1982) to appear.
[13] E.L. Lawler, Combinatorial Optimization: Networks and Matroids (Holt, Rinehart and Winston, New York, 1976).
[14] E.L. Lawler, J.K. Lenstra and A.H.G. Rinnooy Kan, Recent developments in deterministic sequencing and scheduling, presented at the Advanced Study and Research Institute on Deterministic and Stochastic Scheduling, Durham, England, 1981.
[15] R.H. Möhring, On the distribution of locally undecomposable relations and independence systems, Methods of Oper. Res. 42 (1981) 33–48.
[16] R.H. Möhring, Minimizing costs of resource requirements subject to a fixed completion time in project networks, submitted.
[17] R.H. Möhring, Almost all comparability graphs are UPO, submitted.
[18] F.J. Radermacher, Kapazitätsoptimierung in Netzplänen, Math. Syst. in Econ. 40 (Anton Hain, Meisenheim am Glan, 1978).
[19] F.J. Radermacher, Cost-dependent essential systems of ES-strategies for stochastic scheduling problems, Methods of Oper. Res. 42 (1981) 17–31.
[20] F.J. Radermacher, Optimization of resource constraint project networks, submitted.
[21] F.J. Radermacher, Schedule-induced posets, submitted.
[22] A.H.G. Rinnooy Kan, Machine Scheduling Problems: Classification, Complexity and Computations (Nijhoff, The Hague, 1976).
[23] J. Valdes, R.E. Tarjan and E.L. Lawler, The recognition of series-parallel digraphs, Proc. 11th. Annual ACM Symp. on Theory of Comp., ACM (1979) 1–12.

This Page Intentionally Left Blank

Annals of Discrete Mathematics 16 (1982) 241–260

EAR DECOMPOSITIONS OF ELEMENTARY GRAPHS AND GF_2-RANK OF PERFECT MATCHINGS

Denis J. NADDEF*
I.M.A.G., Université Scientifique et Médicale, F-38041 Grenoble Cedex, France

William R. PULLEYBLANK**
Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

We study the rank over GF_2 of the set of incidence vectors of the perfect matchings of a graph, and in particular, the relationship between this value and the number of steps in an ear decomposition of the graph. We show that if a graph has an ear decomposition which is sufficiently simple, the GF_2-rank will be one greater than the number of steps in the ear decomposition. We also give examples which show that this relationship does not remain true in general.

We show that for a special class of 3-connected planar bicritical graphs, the so-called Halin graphs, there does exist such a simple ear decomposition and so the GF_2-rank of the incidence vectors of the perfect matchings can be easily calculated.

1. Introduction

Let $G = (V, E)$ be a graph, where V is the set of vertices and E is the set of edges. For convenience we will assume that G is simple, i.e., without loops or multiple edges. In general we adopt the terminology of Berge [1]. A *perfect matching* or *1-factor* of G is a subset M of E such that every vertex $v \in V$ is incident with exactly one element of M.

We can associate with every subset X of E a (0, 1)-vector $\boldsymbol{X} = (x_e : e \in E) \in \mathbb{R}^{|E|}$ such that $e \in E$ if and only if $x_e = 1$. Such a vector is called the *incidence* (or *representative*) *vector* of X.

Let P be the set of the incidence vectors of all perfect matchings of G and let K be any field. We define the *K-rank* of the perfect matchings of G, denoted by $r_k(G)$, to be the maximum number of K-linearly independent elements of P, where, of course 0 and 1 are taken as the additive and multiplicative identities respectively of K.

* Research supported by Deutscher Akademischer Austauschdienst (DAAD) and Sonderforschungsbereich 21 (DFG), when on leave at Universität Bonn, Institut für Operations Research.

** Supprted by the Universität Bonn, Sonderforschungsbereich 21 (DFG), Institut für Operations Research. On leave from Department of Computer Science, The University of Calgary, Canada.

The case $K = \mathbb{R}$ or equivalently $K = \mathbb{Q}$ was studied in [8]. Sometimes $r_{\mathbb{R}}(G)$ gives a good lower bound on the number $\Phi(G)$ of perfect matchings. Here we concentrate on the case $K = G\mathbb{F}_2$, the field with 0 and 1 as its only elements. Note that we always have $r_{G\mathbb{F}_2}(G) \leqslant r_{\mathbb{R}}(G) \leqslant \Phi(G)$.

A main interest in the latter case lies in the possible relation between this $G\mathbb{F}_2$-rank and the ear decomposition theory of Lovász [5] (see also [6]) for connected graphs such that every edge belongs to a perfect matching.

In Section 2 we describe the ear decomposition and several of its properties. In Section 3 we present a class of graphs for which the ear decomposition gives the $G\mathbb{F}_2$-rank and then in Section 4 we see how this gives us the K-rank (for $K = G\mathbb{F}_2$ or $\mathbb{R}$) of Halin graphs. The last part, Section 5, will concern various related topics.

We conclude this section with some notation and definitions.

Any edge $e \in E$ which belongs to no perfect matching can be deleted from G without affecting the rank of the perfect matchings. Similarly, if an edge e joining verices u and v belongs to every perfect matching, then we can delete e, u, v and all edges incident with u or v without changing the rank. (These operations may result in a reduced graph $G' = (V', E')$, having $E' = \emptyset$. If $V' \neq \emptyset$, then G has no perfect matching so $r_K(G') = r_K(G) = 0$. If $V' = \emptyset$, then G has exactly one perfect matching, and since $M \equiv \emptyset$ is a perfect matching of G', we have $r_K(G') = r_K(G) = 1$.)

Any edge $e \in E$ which belongs to some but not all perfect matchings is called *useful*. We say that G is a U-graph if every edge is useful. (This is closely related to Lovász's definition [5] of an elementary graph—a graph for which the useful edges form a connected spanning subgraph.)

Henceforth, we only consider connected U-graphs. For a nonconnected U-graph G having components $G_1, G_2, \ldots, G_p$, the results can be applied to each component and then $r_K(G) = \Sigma_{i=1}^{p} r_K(G_i) - p$.

Also G will usually be explicitly assigned a particular perfect matching M_0 which we will call the *reference perfect matching*. An *alternating cycle* in G (relative to M_0) is an even elementary cycle such that the edges alternately belong and do not belong to M_0. An *alternating path* is a simple path whose edges alternately belong and do not belong to M_0. An *augmenting path* between a and b is an alternating path of length $2r+1$ ($r \geqslant 0$) linking a and b such that the 2nd, 4th, $\ldots$, $2r$th edges are in M_0; a *decrementing path* between a and b is defined in the same way but with the 1rst, 3rd, $\ldots$, $(2r+1)$th edge in M_0. Note that since M_0 is a perfect matching, our definition of an augmenting path is slightly different from the standard one which requires a and b not to be incident with elements of M_0.

Note that if G is a U-graph and M_0 is a perfect matching of G, then every edge belongs to an alternating cycle. Moreover, a connected U-graph can have no cutnode.

2. Structural properties of U-graphs and the ear decomposition

Throughout this section, $G = (V, E)$ will be a connected U-graph and M_0 will be a perfect matching of G. *Matching labeling* is the following process, which iteratively labels the vertices after one (or more) has been assigned an initial label, and continued until no further operations can be performed:

- If j does not have a label '+', but is joined by an edge of M_0 to a node i with label '−', then label j with '+'.
- If j does not have a label '−', but is joined by an edge of $E - M_0$ to a node i with label '+', then label j with '−'.

(Note that every labeled node has a natural predecessor, the node that caused it to receive a label.)

- As soon as a vertex gets both labels '+' and '−', traceback the two paths of predecessors which led to that double labeling to the first common vertex v. The two paths then form an odd cycle which is shrunk to a single vertex, called a *pseudo-vertex*; give this vertex the label '+'. The old predecessor of v is taken to be the predecessor of the pseudo-vertex. Moreover, every vertex of G contained in the pseudo-vertex is given the single label '+'.

Suppose we initialize by labeling a vertex v with '−'. It is shown in [8] that, when we start with a U-graph, in the resulting graph G^*, every vertex gets a label and the label does not depend on the choice of M_0 or the order of the labeling. This induces a labeling of the vertices of G which can be considered as a function $\lambda_v : V \to \{+, -\}$.

During the labeling process starting from a vertex v, a vertex first receives a label when an alternating path from v, whose first edge is in M_0, is discovered. Thus for a U-graph there is an alternating path starting with an edge of M_0, from v to every other vertex. Moreover, a vertex receives the label '+' when a decrementing path is discovered. (The reader is encouraged to verify these assertions for labelings involving pseudo-nodes!) This matching labeling process is basically the tree growing process of Edmonds' matching algorithm [2].

Now consider the following two relations:

$v \sim w$ if and only if $\lambda_v w = \text{'}-\text{'}\ (=\lambda_v v)$,

$v \approx w$ if and only if there is no decrementing path (relative to M_0) between v and w.

Proposition 2.1. *Relations '$\sim$' and '$\approx$' are the same equivalence relation.*

Proof. First we show that '$\approx$' is an equivalence relation by showing that transitivity is satisfied. If u, v, $w \in V$ and $u \approx w$, then there exists a decrementing path from u to w. Since there exists an alternating path from v to w, which starts with an edge of M_0, we can construct a decrementing path from v to one of u or w, so $u \not\approx v$ or $v \not\approx w$ and we are done.

If there exists a decrementing path from u to v, then a labeling starting with u labeled '$-$' which follows this path gives v the label '+', so $u \not\sim v$. If no such path exists, then at the end of the labeling y must have the label '$-$', so $u \sim v$. Thus the relations '$\sim$' and '$\approx$' are identical. □

The classes of these relations are called the *M-classes* of G.

It is an easy consequence of Berge's alternating path theorem (see [1]) that the following relation:

$$u \cong v \text{ if and only if } G - u - v \text{ has no perfect matching}$$

gives the same classes as '$\approx$'. A consequence is that the M-classes do not depend on the matching M_0 chosen.

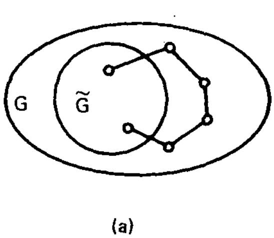
Another characterization of the M-classes [4, 5] is given by the following statement:

> $S \subseteq V$ is a M-class if and only if $G - S$ consists of $|S|$ critical components.

A *critical* graph is a graph with an odd number of vertices such that the deletion of any one leaves a graph containing a perfect matching.

A graph such that every M-class contains a single vertex is said to be *bicritical.* These graphs have the property that deleting any pair of vertices yields a graph which still has a perfect matching.

An *ear* is a path containing an odd number of edges. It may simply consist of a path of length one, that is, two nodes joined by an edge. We say that $G = (V, E)$ is obtained from $\tilde{G} = (\tilde{V}, \tilde{E})$ by an *1-ear addition* if G is obtained from $\tilde{G}$ by joining two vertices of $\tilde{V}$ by an ear, such that only the two end vertices belong to $\tilde{V}$ (see Fig. 1(a)).



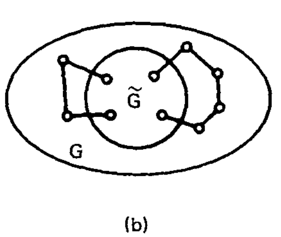


Fig. 1. (a) 1-ear addition; (b) 2-ear addition.

Similarly, we say that $G = (V, E)$ is obtained from $\tilde{G} = (\tilde{V}, \tilde{E})$ by a *2-ear addition* if G is obtained from $\tilde{G}$ by adding two vertex-disjoint ears, which join disjoint pairs of vertices (Fig. 1(b)) of $\tilde{V}$ and for which only these end vertices belong to $\tilde{V}$.

Note that two consecutive 1-ear additions need not in general give a 2-ear addition.

2.1. *The ear decomposition of U-graphs*

An *ear decomposition* of a U-graph G is a sequence $G_0, G_1, \ldots, G_t = G$ of graphs such that

(i) G_0 consists of two vertices joined by an edge,

(ii) $G_1, G_2, \ldots, G_t$ are U-graphs,

(iii) for $i = 1, 2, \ldots, t$, G_i is obtained from G_{i-1} by either an 1-ear addition or a 2-ear addition. In the latter case adding only one of the two ears would not yield a U-graph.

It is a consequence of results of Lovász [5] and Lovász and Plummer [6] that such a decomposition exists. The total number of ears used in the construction of G is equal to $\nu(G)$, the cyclomatic number of G. However, the value t, the number of *steps* of the ear decomposition will depend upon the number of times these ears are added in sets of two. We will now give an algorithmic construction of such a decomposition.

Let G be a U-graph and M_0 a perfect matching of G. Let G_0 be the graph induced by any edge of M_0. Suppose, for $i \geqslant 0$, we have built $G_0, G_1, \ldots, G_i$ satisfying (ii) and (iii) and perfectly matched by the matching induced by M_0. If $G_i = G$, then let $t \equiv i$ and terminate. If not, let J be the set of all edges of $E(G) - E(G_i)$ which are incident with one or two vertices of G_i. Give each edge in J a weight equal to the number of incident vertices of G_i, that is, 1 or 2 depending if the edge is in the coboundary of $V(G_i)$ or not. Give each edge not in J a weight of 0. Find a perfect matching $\tilde{M}_{i+1}$ of G for which the sum of the weights is minimum, *but positive.* (Such a perfect matching exists, because G is a U-graph so every edge of J belongs to a perfect matching. A way to find it is given later.) It follows from a nontrivial theorem of Lovász and Plummer [6, Theorem 5.4] that the weight of $\tilde{M}_{i+1}$ will be either two or four. The symmetric difference $M_0 \triangle \tilde{M}_{i+1}$ consists of alternating cycles, only one of which, say Γ_{i+1}, meets edges of J. Set $G_{i+1} \equiv G_i \cup \Gamma_{i+1}$. Then G_{i+1} is a U-graph and (ii) is satisfied. The portion of Γ_{i+1} not in G_i defines one or two ears depending on the weight of $\tilde{M}_{i+1}$. The second part of (iii) follows from the minimality of $\tilde{M}_{i+1}$. Finally, G_{i+1} is perfectly matched by M_0 and we can repeat the process until G is obtained.

The matching $\tilde{M}_{i+1}$ can be found as follows. Delete individually each edge j

of J, together with its two end points and find a minimum weight perfect matching $M(j)$ in the resulting graph. Let $\tilde{M}_{i+1}$ be a member of $\{M(j) \cup \{j\} \mid j \in J\}$ for which the weight is minimum.

If the ear decomposition $G_0, G_1, \ldots, G_t$ is given, then the matching M_0 which induces perfect matchings on every G_i, $i = 0, \ldots, t$ is uniquely determined. However, it appears to be quite difficult to obtain the ear decomposition without having M_0 at hand to guide the selection of ears. For example consider the graph of Fig. 2. Suppose we choose G_0 to be the graph induced by the edge joining nodes 3 and 4. Choosing the ear induced by the nodes (3, 2, 5, 4) leads us to G_1 which is a U-graph (a 4-cycle), but there is no ear decomposition of G that contains G_1.

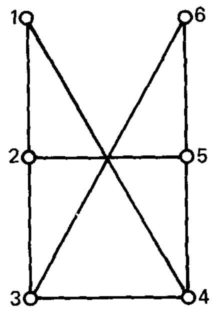


Fig. 2.

We now relate the ear decomposition to the M-classes. When one ear is added to G_i, it links two vertices of G_i of different classes. In fact, joining any two vertices of different M-classes of a U-graph G by an ear yields another U-graph. When a 2-ear addition is performed then each ear joins a pair of vertices and each pair belongs to a different class. Moreover, there exist between the two vertices of the first class and those of the second class two disjoint decrementing paths.

In summary, an 1-ear addition can be performed if and only if the resulting graph has an alternating cycle containing the ear. A 2-ear addition can be performed if and only if the resulting graph has an alternating cycle containing both ears, but no such cycle containing only one.

For $k = 1, \ldots, t$ let $\mathscr{C}_k \equiv \{\Gamma_1, \Gamma_2, \ldots, \Gamma_k\}$ where the Γ_i's are the alternating cycles defined in the algorithm. Let $M_i \equiv M_0 \triangle \Gamma_i$ for all $i = 1, \ldots, t$, and $\mathscr{M}_k \equiv \{M_0, M_1, \ldots, M_k\}$. We now extend each $\mathscr{C}_k$ to a cycle basis $\mathscr{B}_k$ of G_k. Let $\mathscr{B}_1 \equiv \{\Gamma_1\}$ and $\mathscr{B}_k \equiv \mathscr{B}_{k-1} \cup \{\Gamma_k\}$ if G_k was obtained from G_{k-1} by adding one ear, $\mathscr{B}_k \equiv \mathscr{B}_{k-1} \cup \{\Gamma_k, C_p\}$ if G_k was obtained from G_{k-1} by a 2-ear addition, where C_p is a cycle which completes $\mathscr{B}_{k-1} \cup \{\Gamma_k\}$ to a cycle basis. Note that C_p will always contain exactly one of the two ears added at that step.

Theorem 2.2. *Let $G_0, G_1, \ldots, G_t$ be an ear decomposition of G. Then the following four properties are equivalent, where all spaces are taken over $G\mathbb{F}_2$.*

(1) *The set $\mathcal{M}_t$ forms a basis of the perfect matchings of G.*

(2) *$\mathcal{C}_t$ is a basis of the space generated by the alternating cycles of G.*

(3) *In any basis of cycles of G not more than t can be alternating cycles.*

(4) $r_{G\mathbb{F}_2}(G) = t + 1$.

Proof. Clearly $\mathcal{M}_t$ is a linearly independent set of $t+1$ perfect matchings of G and hence (1) and (4) are equivalent. Similarly, $\mathcal{C}_t$ is a linearly independent set of t alternating cycles of G and hence (2) and (3) are equivalent. Now we show that (1) and (2) are equivalent, which completes the proof.

Let Γ be an alternating cycle of G and let $M \equiv \Gamma \triangle M_0$. By (1) there exist $\alpha_0, \ldots, \alpha_t \in G\mathbb{F}_2$, not all equal to zero, such that

$$\boldsymbol{M} = \sum_{i=0}^{t} \alpha_i \boldsymbol{M}_i .$$

(Recall that $\boldsymbol{M}$ denotes the incidence vector of M.) Note that $\Sigma_{i=0}^{t} \alpha_i = 1$ because, if not, the subgraph induced by the symmetric difference of the M_i for which $\alpha_i = 1$ would have even degree.

Since $M_i = M_0 \triangle \Gamma_i$ we have

$$\boldsymbol{M} + \boldsymbol{\Gamma} = \sum_{i=0}^{t} \alpha_i (\boldsymbol{M}_0 + \boldsymbol{\Gamma}_i) \quad \text{where we take } \Gamma_0 \equiv \emptyset.$$

This gives

$$\boldsymbol{M}_0 + \Big(\sum_{i=0}^{t} \alpha_i\Big)\boldsymbol{M}_0 = \sum_{i=1}^{t} \alpha_i \boldsymbol{\Gamma}_i + \boldsymbol{\Gamma} .$$

Since $\Sigma_{i=0}^{t} \alpha_i = 1$ we get

$$\sum_{i=1}^{t} \alpha_i \boldsymbol{\Gamma}_i + \boldsymbol{\Gamma} = 0 ,$$

so Γ is generated by $\mathcal{C}_t$.

Conversely suppose $\mathcal{C}_t$ is a basis of the alternating cycles and let M be any perfect matching of G. Let $\Gamma \equiv M_0 \triangle M$. Then Γ is the union of disjoint alternating cycles so there exist $\alpha_1, \alpha_2, \ldots, \alpha_t$ not all zero such that

$$\boldsymbol{\Gamma} = \sum_{i=1}^{t} \alpha_i \boldsymbol{\Gamma}_i ,$$

and so

$$M_0 + \Gamma = M_0 + \sum_{i=1}^{t} \alpha_i \Gamma_i .$$

Let $\alpha_0 \equiv 1 - \Sigma_{i=1}^{t} \alpha_i$. Then

$$M = \sum_{i=0}^{t} \alpha_i (M_0 + \Gamma_i) \quad \text{(again, setting } \Gamma_0 \equiv \emptyset\text{)},$$

hence

$$M = \sum_{i=0}^{t} \alpha_i M_i ,$$

which shows that $\mathscr{M}_t$ is a basis of the perfect matchings. □

The question which naturally arises now is whether or not every ear decomposition of $G = (V, E)$ satisfies (1) to (4). If such were the case, then the following would also hold:

(5) Given a U-graph $G = (V, E)$, the number of times a 2-ear addition is performed is the same in any ear decomposition of G.

This is because, as noted earlier, the cyclomatic number of G_i, denoted by $\nu(G_i)$, is one or two larger than that of G_{i-1}, depending on whether an 1-ear or a 2-ear addition was performed. Trivially $\nu(G_0) = 0$ and t equals, therefore, $\nu(G)$ minus the number of times a 2-ear addition was performed. If (4) is valid for every ear decomposition, then t is independent of the decomposition, which implies that the number of times 2-ear additions are performed is also independent of the decomposition.

We will now show that these properties need not be true in all cases. Curiously enough, however they do hold for a great many graphs, including the Petersen graph ! In the next part we will show that if a graph G has a 'sufficiently simple' ear decomposition, then we will have $r_{GF_2}(G) = t + 1$. This will then partially answer a question first posed by Lovász who asked whether or not (4) and (5) were always satisfied.

First the reader can convince himself that there is only one ear decomposition of the Petersen graph (up to isomorphism). Necessarily G_1 is an 8-cycle, G_2 is obtained by a 2-ear addition giving a spanning subgraph of Petersen, G_3 is obtained by another 2-ear addition and finally G_4 = Petersen is obtained by adding the last missing edge. So (5) is valid in the case of the Petersen graph. Now the Petersen graph is bicritical so we can continue to build K_{10}, the complete graph on 10 vertices, by 1-ear additions. Thus we obtain an ear decomposition of K_{10} that has two 2-ear additions.

But K_{10} has an ear decomposition with a single 2-ear addition. To find such a decomposition proceed as follows: First perform 1-ear additions until $K_{5,5}$, the complete bipartite graph with 5 vertices on each side, is constructed. Then perform a 2-ear addition, adding an edge between the members of two pairs of vertices on opposite sides of $K_{5,5}$. The resulting graph is bicritical. Then K_{10} can be obtained performing only 1-ear additions. So K_{10} shows us that not every ear decomposition satisfies (1) to (5). In fact the Petersen graph with any edge added to it is also a counterexample.

In general, ear decompositions seem to allow a large amount of variation in the steps performed. In particular, examples can be found which show that the number of 2-ear additions may depend on

- the choice of G_1 (see Petersen graph plus 1 edge),
- the choice of the two ears, when more than one 2-ear addition is possible,
- the decision to perform a 2-ear addition while there is still an 1-ear addition possible.

This motivates the definition of an *optimum ear decomposition* of G as any ear decomposition of G with the minimum number of 2-ear additions, or equivalently with the highest number of steps. We now make the following conjecture.

Conjecture 2.3. *The $G\mathbb{F}_2$-rank of the perfect matchings of a connected U-graph is one greater than the number of steps of an optimum ear decomposition of G.*

3. Optimum ear decompositions

In this section we show that all ear decompositions which are sufficiently simple are, in fact, optimum.

Theorem 3.1. *Let G be a U-graph for which there exists an ear decomposition such that*

(i) *at most two 2-ear additions are performed,*

(ii) *if two 2-ear additions are performed, then the second such addition is the last step of the construction of G.*

Let t be the number of steps in the ear decomposition. Then (1) *to* (4) *hold, and in particular,* $r_{G\mathbb{F}_2}(G) = t + 1$.

Proof. If no 2-ear additions are performed, then $t = \nu(G)$ so (3) holds trivially. So suppose that at least one 2-ear addition is performed.

If a 2-ear addition has been performed, then necessarily the graph is nonbipartite. Any cycle basis of a nonbipartite graph must contain at least one odd cycle. Thus if only one 2-ear addition has been performed, $t = \nu(G) - 1$ so again (3) holds trivially.

Suppose we now perform a second 2-ear addition. The alternating cycle Γ_t contains both ears. If an alternating cycle Γ is not a linear combination of alternating cycles of $\mathscr{C}_t$, it must contain both ears of the last step. For if it contained none of these ears, it would be contained in G_{t-1}, which we just saw is impossible. If it contains only one ear, then the 2-ear addition was not allowable, since an 1-ear addition was possible. So $\Gamma \triangle \Gamma_t$ is a linear combination of cycles of $\mathscr{B}_{t-1}$. By parity, the coefficient of C_1 in that combination is zero, so Γ is a linear combination of alternating cycles, which contradicts our choice of Γ. Hence, again (3) is satisfied. □

Corollary 3.2. *Any ear decomposition satisfying the conditions of Theorem* 3.1 *is an optimum ear decomposition.*

In Fig. 3 an example, which is obtained from the Petersen graph by moving one edge, shows that if an single 1-ear addition is performed after the second 2-ear addition, then this ear decomposition may not be optimum. The graph G of Fig. 3 has at least two ear decompositions. One starts with a hamiltonian cycle of G and uses a single 2-ear addition and so is optimum. The second starts with an 8-cycle and uses two 2-ear additions and ends with a single 1-ear addition after the last 2-ear addition.

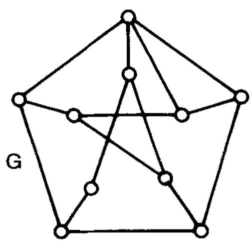


Fig. 3.

Corollary 3.3. *If G has an ear decomposition with a single* 2*-ear addition, then* $r_{GF_2}(G) = r_{R}(G) = \nu(G)$.

Proof. We always have $r_{GF_2}(G) \leq r_R(G)$. Since G must be nonbipartite, $r_R(G) \leq \nu(G)$. This follows either from Theorem 4.5 presented in the next section, or simply from the observation that the degree constraints for a connected nonbipartite graph form a linearly independent set of $|V|$ equations satisfied by all perfect matchings of G.

The corollary now follows from Theorem 3.1. □

Recall that if G is a bipartite U-graph, then every ear decomposition requires no 2-ear addition and $r_{\mathrm{R}}(G) = r_{\mathrm{GF}_2}(G) = \nu(G) + 1$.

4. Matching rank of Halin graphs

A graph G is a Halin graph if it can be constructed as follows:

Let T be a tree having at least four vertices in which each nonleaf has degree at least three. Embed T in the plane and construct a cycle C passing through all leaves of T in such a way that $G = T \cup C$ remains planar (see [3, 7, 9]). Note that Halin graphs can never have loops or multiple edges.

Now let $G = T \cup C$ be a Halin graph. We will always let T^* denote the tree obtained by deleting (or pruning) all leaves of T. Thus T^* is never empty, it always contains at least one node. If it contains exactly one node, which must be adjacent to every node of C, then G is a *wheel* (see Fig. 6).

A leaf of T^* can only be adjacent to one node not in C. Moreover, the adjacent nodes of C must form a consecutive segment of C. For any leaf v of T^*, we define the *fan centered* at v to be the subraph of G induced by v and the adjacent nodes on C. For example, in Fig. 4, u_1, u_2, u_3 are the nodes of a fan centered at u_1 and v_1, v_2, v_3, v_4, are the nodes of a fan centered at v_1. Note that w does not appear in a fan because w is not a leaf of T^*.

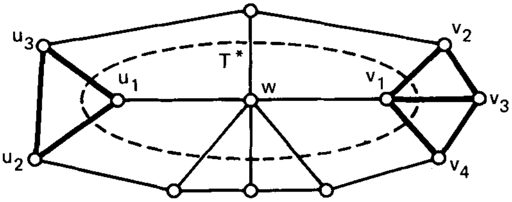


Fig. 4.

Proposition 4.1. *A Halin graph G contains at least two fans unless G is a wheel.*

Proof. If T^* consists of a single node, then H is a wheel. Otherwise, T^* contains at least two leaves and hence G contains two fans. □

It follows from Propositon 4.1 that every Halin graph contains a triangle and hence is nonbipartite. Moreover, it is shown in [7], that these graphs are bicritical when the number of vertices is even, which implies that they are U-graphs.

We now state the main theorem of this section.

Theorem 4.2. *If G is a Halin graph with an even number of vertices, then there exists an ear decomposition requiring only one 2-ear addition. Hence* $r_{GF_2}(G) = r_{\mathbb{R}}(G) = \nu(G)$.

Proof. We are going to construct such a decomposition. Again, T^* is the subtree of T obtained by pruning all leaves. Let M^* be a maximum matching of T^* which saturates all nonleaves of T^*. It can be seen that one leaf of T^* must be saturated, unless $E(T^*) = \emptyset$, in which case G is a wheel.

Let C^* be the subgraph of G obtained by deleting all nodes of T^* which are matched by M^*. Thus, C^* consists of C, together with those fans whose centers are not adjacent to an edge of M^*. It is easily verified that C^* has a hamiltonian cycle C^1. If we add every second edge of C^1 to M^*, we obtain a perfect matching M_0 of G. Moreover C^1 is an alternating cycle with respect to M_0 (see Fig. 5).

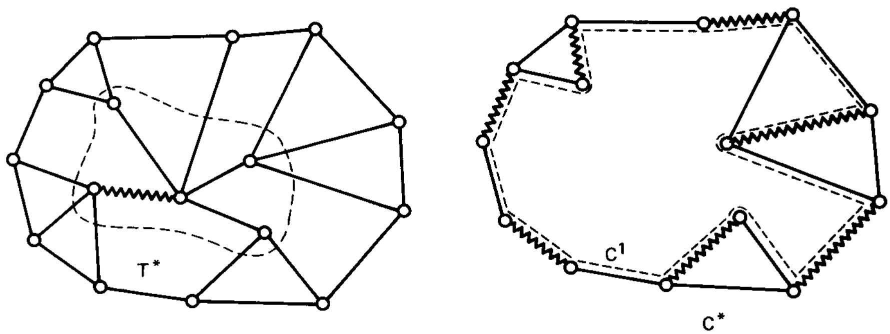


Fig. 5.

We are now going to describe an ear decomposition such that G_1 is the graph induced by C^1, that is $\Gamma_1 = C^1$.

Case 1. G is a wheel, that is, $E(T^*) = \emptyset$ (see Fig. 6).

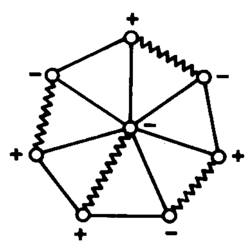


Fig. 6.

M_0 contains one radial edge and all the others are on C. Moreover, $C^* = G$ and C^1 is a hamiltonian cycle of G. Label the vertices with '+' and '−' alternately around C^1. Suppose the center gets the label '−'. We can add every radial edge with ends of opposite label by 1-ear additions. Then there is only one edge with both ends labeled '+' (necessarily on C). Put in that edge together with any radial edge not yet in (which must have both ends labeled '−'). This is a 2-ear addition. Now because of the uniqueness of the edge with both ends labeled '+', every other edge not yet added can be added with a 1-ear addition, which proves Theorem 4.2 for wheels.

Case 2. G is not a wheel, so $E(T^*) \neq \emptyset$ and there is a leaf a of T^* which is saturated by M^*. Let P_a be a maximal alternating path of T^* containing a. Let b be the other end of P_a. Then a is the center of a fan A and b is the center of another fan B. Now G_1 is the graph induced by $\Gamma_1 = C^1$ and so is bipartite; label alternately with '+' and '−' around C^1. This gives the two M-classes which in turn give the bipartition of G_1. Since a was matched by an edge of T^*, a is not in $V(G_1)$, and so is unlabeled, but is adjacent to vertices labeled '+' and others labeled '−'. (Every fan has labels of both kinds.)

Consider now the two following subcases.

Case 2a. b is unlabeled, so $b \notin V(C^1)$ (see Fig. 7).

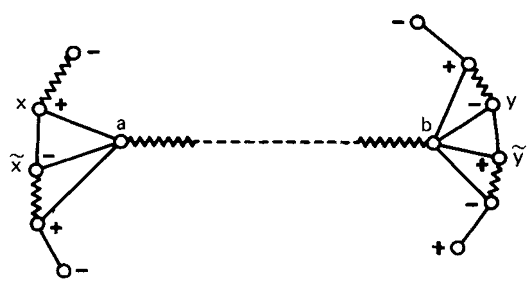


Fig. 7.

Choose any vertex x of the fan A and any vertex y of the fan B of opposite labels. Then the path ε_1 consisting of P_a and edges linking x and a on one side, b and y on the other side is a valid ear. (It links nodes of different M-classes.) Let G_2 be the graph obtained by adding ε_1 to G_1.

Let $\tilde{x}$ be a neighboring vertex of x on C^1 and on fan A, and let $\tilde{y}$ be similarly defined for y in fan B. Instead of ε_1 it would have been possible to add the ear ε_1 consisting of P_a together with the edges between $\tilde{x}$ and a and between b and $\tilde{y}$. (Again, it links node of different M-classes.) This shows that these two edges belong to some alternating cycle of G_2, and each by itself does not complete an alternating cycle so the addition of those two edges is a valid 2-ear addition which leads to G_3.

Case 2b. b is labeled, that is, $b \in V(C^1)$.

Without loss of generality we suppose b is labeled '+'. (See Fig. 8 where C^1 traverses the fan B in the following order: b_1, b_2, b, b_3, b_4.)

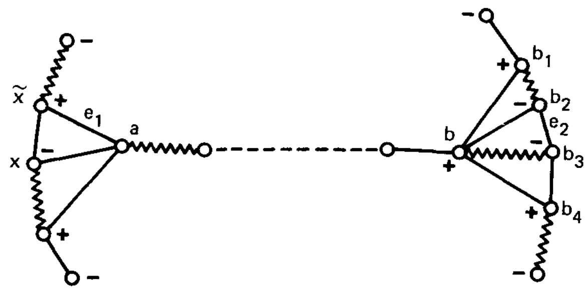


Fig. 8.

Choose in fan A any vertex x of label '$-$' (the opposite of that of b). Then the path ε_1 consisting of P_a together with the edge between a and x is a valid 1-ear addition. Let G_2 be the graph obtained.

Choose any vertex $\tilde{x}$ of fan A adjacent to x and on C, let e_1 be the edge joining a and $\tilde{x}$. Let e_2 be the unique edge of the fan B which is on C but not on C^1. (It has both ends labeled '$-$' in our case.) As a consequence of the following property, (6), we can add e_1 and e_2 as a valid 2-ear addition, giving the graph G_3.

Consider the following subgraph of G consisting of the graph induced by two consecutive radial edges of a fan, the edges of C joining their endpoints on C, and the edges of the coboundary of this triangle (see Fig. 9). The reader can easily verify the following property (see Fig. 10(a), (b) and (c)).

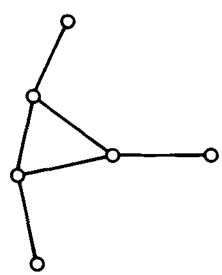


Fig. 9.

(6) Let M be any matching saturating the three nodes of the triangle. An alternating path arriving at the center of the fan can continue as an alternating path either clockwise or anticlockwise on C^1.

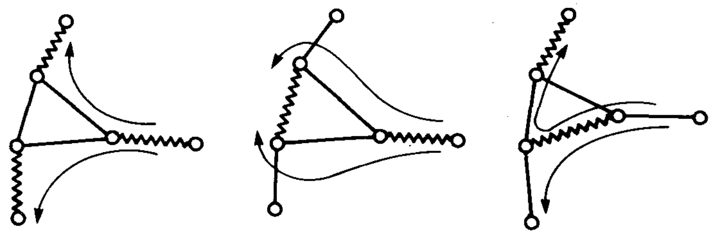


Fig. 10.

This motivates the following definition: A *switch* is any fan such that three edges of it forming a triangle are already present in the current graph G_k of the ear decomposition. In particular if every edge of a fan is in G_k, then it is a switch.

It is easy to see that, once G_3 is constructed, every edge of the fans A and B can be added by a sequence of 1-ear additions.

Let f denote the number of fans of G, or equivalently, the number of leaves of T^*. The next portion of the ear construction will consist of $f-2$ repetitions of the following sequence of 1-ear additions. First an alternating path is added linking a vertex of the portion of T^* already added to a vertex of a previously unadded fan. Then the remaining edges of this fan are added one after the other. We adopt the following notation. For $2 \leqslant i \leqslant f$, G_0^i will denote the graph obtained after i complete fans have been added to the constructed graph. Thus, in particular, we now denote the graph constructed at this point by G_0^2. Our complete ear decomposition, therefore, will be denoted by $(G_1, G_2, G_3, \ldots, G_0^2, G_1^2, \ldots, G_0^3, G_1^3, \ldots, G_0^f, \ldots)$.

We will let T_i^* denote the portion of T^* which exists in G_0^i. Our construction will ensure that G_0^i for $i \geqslant 2$ will have the following properties:

(7) T_i^* is connected.

(8) For any leaf v of T^* which belongs to T_i^*, the entire fan centered at v is in G_0^i.

(9) For any nonleaf v of T^* which belongs to T_i^* there is an alternating path contained in T_i^* linking v to a switch and starting with a matching edge.

Clearly (7)–(9) are satisfied for G_0^2. We now describe the sequence of steps that will obtain G_0^{i+1} from G_0^i for $2 \leqslant i < f$ and which we will see preserves these properties. Then we will describe a sequence of 1-ear additions to complete G_0^f to G.

If there is a part of T^* not in G_0^i, then there is an edge $e \in E(T^*) - E(T_i^*)$, which is incident to a vertex of T_i^*. Let Π_e be a maximal alternating path contained in $T^* - T_i^*$ starting with e. One end vertex w of Π_e is the center of a fan W. Let $\varepsilon_i \equiv \Pi_e$ if w is on C^1, and let $\varepsilon_i \equiv \Pi_e$ plus any edge of the fan W incident with w, if w is not on C^1.

We are going to show that ε_i is a valid ear. That is, adding ε_i to G_0^i will give us a U-graph, which we will denote by G_1^i. By (9), the vertex x of e belonging to T_i^* is linked to a fan by an alternating path Π_e^* starting with a matching edge. Because T^* is a tree, Π_e and Π_e^* are disjoint, hence so too are ε_i and Π_e^*. If ε_i is forced on C^1 in one direction, by (9) the fan at the end of Π_e^* is a switch and so can send $\varepsilon_i + \Pi_e^*$ in the opposite direction so we get an alternating cycle containing ε_i, so ε_i is a valid ear.

Now the fan W can be entered entirely by a sequence of 1-ear additions giving G_j^i for $j > 1$ until the last such edge is added which gives G_0^{i+1}. Clearly (7)–(9) are still satisfied.

If $G_0^f \neq G$, then the edges of G not in G_0^f join nonleaves of T^* to vertices of C. (See Fig. 11, in this case G_0^f is everything except edges g to 1.) Let $\bar{e}$ be such an edge.

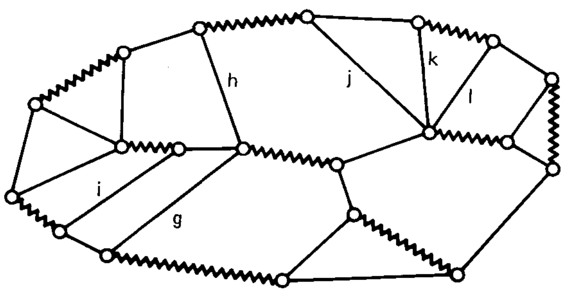


Fig. 11.

The path consisting of $\bar{e}$ and its two adjacent vertices is forced on C^1 in a certain direction, and by (9), is linked on the other side to a fan which is a switch. So $\bar{e}$ belongs to an alternating cycle and hence is a valid ear. Therefore, every edge of T not yet added can be added by an 1-ear addition. This completes the proof, the equalities of Theorem 4.2 being a consequence of Corollary 3.3 of Theorem 3.1. □

For a Halin graph $G = T \cup C$, the cyclomatic number equals $|V(C)|$ so we have the following corollary.

Corollary 4.3. *If $G = T \cup C$ is a Halin graph with an even number of nodes, then* $r_{GF_2}(G) = r_{\mathbb{R}}(G) = \nu(G) = |V(C)|$.

The result that $r_{\mathbb{R}}(G) = |V(C)|$ was first proved in [9], using a very different approach. The idea of this proof was to use the polyhedral properties of dimension, a characterization of a minimal linear system necessary to define the convex hull of the incidence vectors of perfect matchings of a graph and the following result. We say that $S \subseteq V$ is a *tight set* if $G(S)$ (the subgraph of G induced by S) is critical, and every perfect matching of G contains exactly one edge of $\delta(S)$, the coboundary of S. A tight set is *trivial* if $\delta(S) = \delta(i)$ for some node $i \in V$.

Theorem 4.4 [9, Theorem 2.8]. *A Halin graph with an even number of nodes has no nontrivial tight set.*

We now show how this can be derived from Theorem 4.2 and a result in [8] on $r_{\mathrm{R}}(G)$.

We say that a family $\mathcal{S}$ of subsets of $V(G)$ is *nested* if for any $S, T \in \mathcal{S}$, either $S \cap T = \emptyset$ or $S \supseteq T$ or $T \subseteq S$. If $\mathcal{S}$ is a nested family of subsets of $V(G)$, then we define $G \times \mathcal{S}$ to be the graph obtained from G by contracting each of the (necessarily disjoint) maximal members of $\mathcal{S}$ to form a new vertex. For any $T \in \mathcal{S}$ we let $\mathcal{S}[T] = \{S \in \mathcal{S}: S \not\subseteq T\}$. We say that a nested family $\mathcal{S}$ has the *odd cycle property* if $G(T) \times \mathcal{S}[T]$ is nonbipartite for every $T \in \mathcal{S}$.

Now for any graph G, we let $\mathcal{T}(G)$ denote the set of $S \subseteq V$ such that S is a tight set and $|S| > 1$. Note that if G is bipartite, then it contains no critical subgraphs, and hence $\mathcal{T}(G) = \emptyset$. Conversely, it can be shown (see [8]) that a nonbipartite U-graph contains a tight set S with $|S| > 1$. Finally, note that $\mathcal{T}(G)$ may contain some trivial tight sets. For any vertex v such that $G - v$ is critical, $V - \{v\}$ is a trivial tight set belonging to $\mathcal{T}(G)$.

Theorem 4.5 [8, Theorem 13]. *Let G be a connected nonbipartite U-graph. Let $\mathcal{S}$ be a nested subfamily of $\mathcal{T}(G)$ having the odd cycle property and which is maximal by inclusion. Then*

$$r_{\mathrm{R}}(G) = |E(G)| - |V(G)| - |\mathcal{S}| + 2 = \nu(G) - |\mathcal{S}| + 1 .$$

(If G is not connected, then the '2' in the formula for $r_{\mathrm{R}}(G)$ should be replaced with '$p + 1$' where p is the number of components.)

Now we show why Theorem 4.2 and Theorem 4.5 imply Theorem 4.4.

Let G be a Halin graph and let $\mathcal{S}$ be as in Theorem 4.5. Theorems 4.2 and 4.5 imply $|\mathcal{S}| = 1$. Let $\mathcal{S} = \{S\}$. Suppose S is nontrivial. Let $\tilde{G} \equiv G \times S$, the graph obtained by shrinking S, and let τ be the node of $\tilde{G}$ representing S. It is easy to see that $\tilde{G}$ is a U-graph. If $\tilde{G}$ is nonbipartite, then, as we noted, there is a set $\tilde{S}$ of vertices of $\tilde{G}$ which is tight in $\tilde{G}$ ($|\tilde{S}| \geqslant 3$). If $\tau \in \tilde{S}$, then let

$S' \equiv \tilde{S} \cup S - \{\tau\}$ and if $\tau \not\in \tilde{S}$, then let $S' \equiv \tilde{S}$. Then S' is tight in G, distinct from S and either contains S or disjoint from S. Hence $\{S, S'\}$ is a nested subfamily of $T(G)$ having the odd cycle property, which contradicts the maximality of $\mathscr{S}$. Therefore $\tilde{G}$ is bipartite and has the same number of vertices on each side. Since we assume S to be nontrivial, this number is at least two. Let u and v be two vertices of $\tilde{G}$ belonging to the side that does not contain τ. Then it is easily seen that $G-u-v$ has no perfect matching which contradicts the fact that a Halin graph is bicritical.

5. Some questions about $G\mathbb{F}_2$-rank and ear decompositions

A main question which arises concerning the $G\mathbb{F}_2$-rank of matchings is how large the difference can be between the $\mathbb{R}$-rank and the $G\mathbb{F}_2$-rank. Of course if we do not restrict ourselves to connected graphs, the difference can be as large as we want.

Question 1. Is it true that for connected U-graphs, the difference is at most one ? (Note that of necessity, a connected U-graph is nonseparable.)

A related question is how many 2-ear additions will be performed in a optimum ear decomposition ? If we do not restrict ourselves to 3-connected U-graphs we can build examples where the number of those additions is any desired number. The example shown in Fig. 12, where there are k triangles, requires k 2-ear additions. However it contains a great many vertex cutsets each of cardinality two.

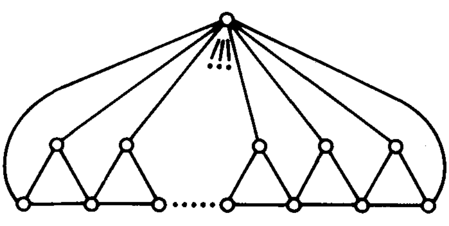


Fig. 12.

Question 2. Let G be a 3-connected U-graph. Is it true that an optimum ear decomposition of G requires at most two 2-ear additions ?

Question 2'. If the answer to Question 2 is negative, is it true when G is moreover bicritical ?

A positive answer to Question 2′, together with Theorem 4.5, the fact that $r_{GF_2}(G) \leqslant r_{\mathbb{R}}(G)$ and a result of Lovaśz–Pulleyblank which states that 3-connected bicritical graphs have no nontrivial tight set, would give a positive answer to Question 1 for those graphs.

A question related to the preceding one is the following.

Question 3. Can we characterize those graphs which have a good ear decomposition with only one 2-ear addition, two 2-ear additions, etc. ?

An answer to the twice 2-ear case cannot be excluding homeomorphic subgraphs of Petersen because we saw that the graph consisting of Petersen with an extra edge only requires one 2-ear addition.

Edmonds and Lovász (see [8]) developed a polynomial algorithm which finds a $\mathbb{R}$-basis of the set of all perfect matchings of a graph. This, of course, determines the $\mathbb{R}$-rank. We ask ourselves the following question.

Question 4. Does there exist a polynomial algorithm to find a $G\mathbb{F}_2$-basis of the set of all perfect matchings of a graph ?

Question 4′. Does there exist a polynomial algorithm to compute the $G\mathbb{F}_2$-rank of a graph ?

Note that it is possible for the answer to Question 4′ to be 'yes', while the answer to Question 4 is 'no'.

If we translate the Edmonds–Lovász procedure from $\mathbb{R}$ to $G\mathbb{F}_2$, then we can see that Question 4 is equivalent to the following.

Question 5. Does there exist a polynomial procedure for finding, if one exists, a perfect matching of $G = (V, E)$ which contains an odd number of edges from a specified subset $\tilde{E}$ of E ?

Acknowledgement

This article has benefited from the helpful comments and suggestions of Michael Plummer and François Jaeger, whose contributions are greatly appreciated.

References

[1] C. Berge, Graphes et Hypergraphes (Dunod, Paris, 1973).
[2] J. Edmonds, Paths, trees and flowers, Canad. J. Math. 17 (1965) 449–467.
[3] R. Halin, Studies on minimally n-connected graphs, in: D.J.A. Welsh, ed., Combinatorial Mathematics and its Applications (Academic Press, New York, 1971) pp. 129–136.
[4] A. Kotzig, Ein Beitrag zur Theorie der endlichen Graphen mit linearen Faktoren I, II, II, Mat. Fyz. Casopis 9 (1959) 73–91, 136–159; id. 10 (1960) 205–215.
[5] L. Lovász, On the structure of factorizable graphs, Acta Math. Acad. Sci. Hungar. 23(1/2) (1972) 179–195.
[6] L. Lovász and M. Plummer, On bicritical graphs, in: Infinite and Finite Sets, Colloq. Math. Soc. János Bolyai 10 (1973) 1051–1079.
[7] L. Lovász and M. Plummer, On a family of planar bicritical graphs, Proc. London Math. Soc. 30(3) (1975) 168–176.
[8] D. Naddef, Rank of maximum matchings in a graph, Math. Programming 22 (1982) 52–70.
[9] W. Pulleyblank, The matching rank of Halin graphs, Rept. No. 80-165, Institut für Ökonometrie and Operations Research, Universität Bonn (1980).

Annals of Discrete Mathematics 16 (1982) 261–280

MIN–MAX RELATIONS FOR DIRECTED GRAPHS*

A. SCHRIJVER

Instituut voor Actuariaat en Econometrie, Universiteit van Amsterdam, Jodenbreestraat 23, Amsterdam, Holland

We prove the following. Let $D = (V, A)$ and $D' = (V, A')$ be directed graphs, both with vertex set V, where D' is acyclic such that each pair of source and sink of D' is connected by a directed path in D'. Suppose that each nonempty proper subset of V which is not entered by any arrow of D', is entered by at least k arrows of D. Then A can be split into classes $A_1, \ldots, A_k$ such that the directed graph $(V, A' \cup A_i)$ is strongly connected, for each i.

This theorem contains as special cases Menger's theorem, Gupta's theorem, Edmonds' branching theorem, a 'bi-branching theorem', a special case of a conjecture of Edmonds and Giles, and a theorem of Frank. The proof yields a polynomial algorithm for finding the splitting as required.

Besides, a slight extension of the Lucchesi–Younger theorem is given.

0. Introduction

Let $D = (V, A)$ and $D' = (V, A')$ be directed graphs, both with vertex set V. Call a subset A'' of A a *strong connector* (*for* D') if the directed graph $(V, A' \cup A'')$ is strongly connected. If V' is a nonempty proper subset of V such that no arrow of D' enters V', the set of arrows of D entering V' is called a *strong cut* (*induced by* D').

We prove the following theorem.

If D' is acyclic and each pair of source and sink of D' is connected by a directed path in D', then the maximum number of pairwise disjoint strong connectors for D' is equal to the minimum size of a strong cut induced by D'. (0.1)

This min–max relation has the following corollaries.

(i) *Menger's theorem* [19]. Let r and s be two vertices of the directed graph $D = (V, A)$. If no set with less than k arrows intersects each directed path from r to s, then there are k pairwise arrow-disjoint such paths. This follows from (0.1) by taking $A' = \{(v, w) \mid v = s \text{ or } w = r\}$. A subset A'' of A is a strong connector for $D' = (V, A')$ if and only if A'' contains a path from r to s.

* Research supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

(ii) *Gupta's theorem* [11]. Let $G=(V,E)$ be a bipartite graph of minimum degree k. Then E contains k pairwise disjoint subsets, each covering V. For, if V' and V'' are the two colour classes of G, let D arise from G by orienting all edges of G from V'' to V', and let $A'=\{(v',v'')\mid v'\in V', v''\in V''\}$. Now a subset A'' of A is a strong connector for D' if and only if A'' covers V.

(iii) *Edmonds' branching theorem* [2]. Let $D=(V,A)$ be a directed graph, and let r be a vertex of D. If each nonempty subset of $V\backslash\{r\}$ is entered by at least k arrows of D, then A contains k pairwise disjoint r-branchings. Here an *r-branching* is a set A'' of arrows such that each vertex of D is reachable by a directed path from r in A''. The result follows from (0.1) by taking $A'=\{(v,r)\mid v\in V\backslash\{r\}\}$. Then A'' is an r-branching if and only if A'' is a strong connector for D'.

(iv) *A bi-branching theorem.* Let $D'=(V,A)$ be a directed graph, and let V be split into classes V' and V''. Suppose each nonempty subset of V' is entered by at least k arrows, and each nonempty subset of V'' is left by at least k arrows. Then A contains k pairwise disjoint bi-branchings. Here a subset A'' of A is called a *bi-branching* (with respect to the splitting V', V'') if each vertex in V' is the end point of some directed path in A'' starting in V'', and each vertex in V'' is the starting point of some directed path in A'' ending in V'. So for $V''=\{r\}$ we obtain r-branchings. The result follows from (0.1) by taking $A'=\{(v',v'')\mid v'\in V', v''\in V''\}$. Then A'' is a bi-branching if and only if A'' is a strong connector for D'.

(v) *A special case of a conjecture of Edmonds and Giles* [3]. Let $D'=(V,A')$ be a directed graph, and let C be a subset of A' such that each directed cut of D' contains at least k arrows of C. (A *directed cut* is the set of arrows entering some nonempty proper subset V' of V, provided that no arrow leaves V'.) Edmonds and Giles conjectured that C can be split into k classes $C_1,\dots,C_k$ such that each C_i intersects each directed cut (i.e., such that contracting the arrows in C_i makes D' strongly connected). Although the general conjecture appeared to be not true (cf. [20]), in the special case that D' is acyclic and each pair of source and sink of D' is connected by a directed path, the conjecture follows from (0.1) by taking A to be the collection of arrows in C with reversed orientation. Then a subset of A is a strong connector for D' if and only if the corresponding subset of C intersects each directed cut of D'. (This special case of the conjecture was announced independently by D.H. Younger.)

(vi) *A theorem of Frank* [5]. Let $D=(V,A)$ be a directed graph, let r be a vertex of D, and let $\mathcal{F}$ be a collection of subsets of $V\backslash\{r\}$ closed under taking unions and intersections. Suppose that each nonempty set in $\mathcal{F}$ is entered by at least k arrows in D. Then A can be split into classes $A_1,\dots,A_k$ such that each nonempty set in $\mathcal{F}$ is entered by at least one arrow in each of the A_i. This follows from (0.1) by taking A' to be the set of all pairs (v,w) which do not enter any set in $\mathscr{F}$. (Possibly D' is made acyclic by contracting strong components.) Actually, Frank proved the more general result where it suffices to require $\mathscr{F}$ to be closed under taking unions and intersections of *intersecting* sets in $\mathscr{F}$.

Remark. The condition of D' being acyclic is not essential. Requiring D' to satisfy the conditions after contracting its strong components is sufficient. Actually, it is not difficult to see that (0.1) is equivalent to: let $D = (V, A)$ be a directed graph, and let $\mathscr{F}$ be a collection of subsets of V closed under taking unions and intersections, such that no V_1, V_2, V_3 in $\mathscr{F}\backslash\{\emptyset, V\}$ have $V_1 \cap V_2 \cap V_3 = \emptyset$ and $V_1 \cup V_2 \cup V_3 = V$. If each set in $\mathscr{F}\backslash\{\emptyset, V\}$ is entered by at least k arrows of D, then A contains k pairwise disjoint sets $A_1, \ldots, A_k$ such that each set in $\mathscr{F}\backslash\{\emptyset, V\}$ is entered by an arrow in each A_i.

The corollaries (i)–(vi) are not independent; one easily derives the following implications: (iv) $\Rightarrow$ (iii) $\Rightarrow$ (i), (vi) $\Rightarrow$ (iii), (iv) $\Rightarrow$ (ii), and (v) $\Rightarrow$ (ii). In fact, our proof essentially shows some more implications.

In Section 1 we first give, for the sake of completeness, a proof of Edmonds' branching theorem (iii), by slightly adapting the proof of Lovász [16]. Second, in Section 2, we prove the following general theorem on pairs of submodular functions. (A function f defined on the subsets of a set X is called *submodular* if $f(X') + f(X'') \geqslant f(X' \cap X'') + f(X' \cup X'')$ for all subsets X' and X'' of X.)

> *Let f_1 and f_2 be integral submodular set-functions on a set X, such that $f_i(X') \geqslant \max\{|X'|, k\}$ for each nonempty subset X' of X, and $i = 1, 2$. Then X can be split into classes $X_1, \ldots, X_k$ such that $f_i(X') \geqslant \sum_{j=1}^{k} \max\{|X' \cap X_j|, 1\}$ for each nonempty subset X' of X and $i = 1, 2$.* (0.2)

Actually, this is a theorem on the splitting of vectors in polymatroids (cf. [1] and the remark in Section 2). It generalizes the edge-colouring theorems of König [13] and Gupta [11] in a similar way as Edmonds' matroid intersection theorem [1] generalizes the König–Egerváry theorem [4, 14] on matchings in bipartite graphs.

Third, in Section 3, we show that (0.2) allows us to glue branchings together to form bi-branchings, and thus to extend (iii) to (iv). In Section 4 we deduce, with some induction arguments, (v) from (iv). Finally, in Section 5, we apply a direct construction to obtain the general Theorem (0.1) from (v). Note that, by replacing arrows by parallel arrows, one easily obtains a 'weighted' version of (0.1).

In Section 6 we use this last 'direct construction' also to observe that the following can be derived from the Lucchesi–Younger theorem [18].

Let $D = (V, A)$ and $D' = (V, A')$ be directed graphs, such that for any arrow (v, w) of D there are vertices v' and w' in V, and directed paths in D' from v to v', from w' to v', and from w' to w (cf. Fig. 1, where the wriggled lines stand for directed paths in D'). Let $l: A \to \mathbb{Z}_+$ be some 'length' function. Then the minimum length of a strong connector for D' is equal to the maximum number of strong cuts induced by D' such that no arrow a is in more than $l(a)$ of these strong cuts. (0.3)

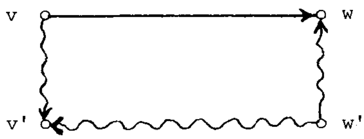


Fig. 1.

If A is the collection of reversed arrows of D', the assumption is obviously satisfied and assertion (0.3) is just the Lucchesi–Younger theorem. If D' is as in (i), (ii), (iii) and (iv) above, we obtain, successively, an (easy) theorem of Fulkerson [7], König's theorem on minimum coverings in a bipartite graph [15], Fulkerson's branching theorem [9], and another 'bi-branching theorem': if the vertex set V of the directed graph $D = (V, A)$ is split into classes V' and V'', and if $c: A \to \mathbb{Z}_+$ is some capacity function, then the minimum capacity of a bi-branching is equal to the maximum number of nonempty proper subsets $V_1, \ldots, V_k$ of V such that $V_i \subset V'$ or $V' \subset V_i$ for each i, and no arrow a of D enters more than $c(a)$ of the V_i.

The conditions for D and D' given in (0.3) are less restrictive than those given in (0.1). In fact, for acyclic D', there is a directed path between each pair of source and sink, if and only if *each* pair (v, w) of vertices of D' is connected by a path of the form of the wriggled lines in Fig. 1. In (0.1) we may not relax the conditions on D' to those given in (0.3), as is shown by the counterexample to the conjecture of Edmonds and Giles (cf. (iv) above). Moreover, if $D = (V, A)$ and $D' = (V, A')$ are as in Fig. 2, where light and heavy lines represent the arrows of D and D', respectively, then any strong connector for D' has cardinality at least 3, whereas any strong cut induced by D' contains at least 2

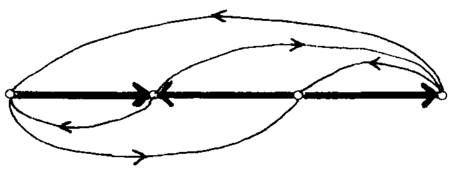


Fig. 2.

arrows of D. Since $|A| = 5$, it is not sufficient to require in (0.1) or (0.3) D' to be weakly connected.

In Section 7 we discuss some generalizations of the results, in terms of sub- and supermodular functions defined on directed graphs, following the lines set out by Edmonds and Giles [3] and Frank [5]. In fact, we give a generalization of (0.3) which slightly extends the theorem of Frank. We also comment on similar extensions of (0.1) and (0.2).

Finally, in Section 8, we formulate the results in terms of polyhedra and linear programming, and this yields, by the ellipsoid method as described in [10], the polynomial solvability of most of the problems. Besides, our proof of (0.1) and (0.2) above will be polynomially constructive (using the fact that the minimum value of a submodular set-function can be found in polynomial time [10]), yielding a polynomial algorithm for optimum packing of strong connectors.

Some terminology. Above we gave already the, rather standard, definitions of submodular function, r-branching and directed cut, and we introduced the notion of bi-branching. A function g is *supermodular* if $-g$ is submodular. We shall sometimes use the easy observation that if f is a submodular, and g is a supermodular set-function on X with $g(X') \leq f(X')$ for all $X' \subset X$, then the collection of sets X' with $g(X') = f(X')$ is closed under taking unions and intersections.

The *indegree* (*outdegree*, respectively) of a set V' of vertices of a directed graph $D = (V, A)$ is the number of arrows of D entering V' (leaving V', respectively), and is denoted by $d_A^-(V')$ ($d_A^+(V')$, respectively).

If c is a rational-valued function defined on a set X, and X' is a subset of X, then, by definition,

$$c(X') := \sum_{x \in X'} c(x) .$$

If c is called a capacity function, then $c(X')$ is the *capacity* of X'.

We note that directed graphs may have multiple arrows, but that we often speak of 'the arrow (v, w)', where 'an arrow from v to w' would be formally more correct.

1. Edmonds' branching theorem

We first give, for the sake of completeness, a proof of a theorem of Edmonds [2], by adapting the method of Lovász [16]. By Edmonds' branching theorem usually is understood the case where $V_1 = \cdots = V_k = \{r\}$.

Theorem 1. *Let $D=(V, A)$ be a directed graph, and let $V_1, \ldots, V_k$ be subsets of V. Suppose $d_A^-(V') \geqslant h(V')$ for each nonempty subset V' of V, where $h(V')$ denotes the number of i with $V' \cap V_i = \emptyset$. Then A can be split into classes $A_1, \ldots, A_k$ such that for each i and each v in V there is a directed path in A_i starting in V_i and ending in v.*

Proof. By induction on $\Sigma_{i=1}^k |V \backslash V_i|$, the case $V_1 = \cdots = V_k = V$ being trivial. Denote by

$$h_{X_1, \ldots, X_t}(V') \tag{1.1}$$

the number of $i = 1, \ldots, t$ with $V' \cap X_i = \emptyset$.

Suppose that $V_1 \neq V$ (say), and consider the collection $\mathscr{F}$ of subsets V' of V with

$$d_A^-(V') = h_{V_2, \ldots, V_k}(V'). \tag{1.2}$$

Note that here the inequality $\geqslant$ always holds, and that (1.2) implies that $V' \cap V_1 \neq \emptyset$. Since the left-hand side of (1.2) is a submodular, and the right-hand side is a supermodular function, the collection $\mathscr{F}$ is closed under unions and intersections. Moreover, $V \in \mathscr{F}$, so there exists a minimal set V' in $\mathscr{F}$ with $V' \not\subseteq V_1$. As

$$d_A^-(V' \backslash V_1) \geqslant h_{V_1, \ldots, V_k}(V' \backslash V_1) > h_{V_2, \ldots, V_k}(V') = d_A^-(V'), \tag{1.3}$$

there is an arrow $a = (v, w)$ from $V' \cap V_1$ to $V' \backslash V_1$. We show that

$$d_{A \backslash a}^-(V'') \geqslant h_{V_1 \cup w, V_2, \ldots, V_k}(V''), \tag{1.4}$$

for each nonempty subset V'' of V. By induction this implies the theorem, as we can split $A \backslash \{a\}$ into classes as required with respect to $V_1 \cup \{w\}, V_2, \ldots, V_k$, and hence, by adding the arrow a to the first class, we obtain a splitting of A as required for $V_1, \ldots, V_k$.

To show (1.4), suppose $V'' \neq \emptyset$ violates (1.4). Since

$$d_A^-(V'') \geqslant h_{V_1, \ldots, V_k}(V'') \geqslant h_{V_1 \cup w, V_2, \ldots, V_k}(V'') > d_{A \backslash a}^-(V'') \geqslant d_A^-(V'') - 1, \tag{1.5}$$

we know that a enters V'', that $w \in V''$, and that

$$d_A^-(V'') = h_{V_1 \cup w, V_2, \ldots, V_k}(V'') = h_{V_2, \ldots, V_k}(V''). \tag{1.6}$$

So V'' is in $\mathcal{F}$, and hence $V' \cap V''$ is in $\mathcal{F}$. Since $V' \cap V'' \not\subset V_1$ as $w \in V' \cap V''$, and since $V' \cap V'' \neq V'$ as $v \notin V''$, this contradicts the minimality of V'. □

2. Pairs of submodular functions

In order to glue branchings together to obtain bi-branchings, we prove a theorem on submodular functions, which has as direct corollaries the theorems of König [13] and Gupta [11] on edge-colourings of bipartite graphs. Also the more general theorem of De Werra [21] may be derived: if (V, E) is a bipartite graph and k is a natural number, then E can be split into classes $E_1, \ldots, E_k$ such that each vertex v is covered by $\min\{d(v), k\}$ of the E_i, where $d(v)$ denotes the degree of v.

Theorem 2. *Let f_1 and f_2 be integral submodular set-functions on a set X, such that*

$$f_i(X') \geq \max\{|X'|, k\} \tag{2.1}$$

for each nonempty subset X' of X, and $i = 1, 2$. Then X can be partitioned into classes $X_1, \ldots, X_k$ such that

$$f_i(X') \geq \sum_{j=1}^{k} \max\{|X_j \cap X'|, 1\} \tag{2.2}$$

for each nonempty subset X' of X, and $i = 1, 2$.

Proof. (i) We first prove the theorem for $k = 2$. Let $Y_1, \ldots, Y_s$ be the minimal nonempty subsets of X with $f_1(Y_j) = |Y_j|$. So the $Y_1, \ldots, Y_s$ are pairwise disjoint, since the collection of sets X' with $f_1(X') = |X'|$ is closed under taking unions and intersections. Moreover, $|Y_j| \geq 2$ for each j, since $f_1(X') \geq 2$ for all nonempty sets X'.

Similarly, let $Z_1, \ldots, Z_t$ be the minimal nonempty subsets of X with $f_2(Z_j) = |Z_j|$. Again, $Z_1, \ldots, Z_t$ are pairwise disjoint and contain at least two elements.

Hence X can be partitioned into classes X_1, X_2 such that both X_1 and X_2 intersect each of $Y_1, \ldots, Y_s, Z_1, \ldots, Z_t$. We prove that (2.2) is satisfied for this choice of X_1 and X_2. Let X' be a nonempty subset of X. If $X_1 \cap X' \neq \emptyset \neq X_2 \cap X'$, then (2.2) follows from (2.1). So we may suppose that $X_2 \cap X' = \emptyset$. Then X' does not contain any of the $Y_1, \ldots, Y_s, Z_1, \ldots, Z_t$, implying that $f_1(X') > |X'|$ and $f_2(X') > |X'|$, which proves (2.2).

(ii) In order to prove the theorem for arbitrary $k \geq 2$, let $X_1, \ldots, X_k$ be pairwise disjoint subsets of X such that (2.2) holds and such that $|X_1 \cup \cdots \cup X_k|$ is as large as possible. If $X_1 \cup \cdots \cup X_k = X$ we are finished, so suppose that $x \in X \setminus (X_1 \cup \cdots \cup X_k)$. Consider the collection $\mathcal{F}$ of all subsets X' of X with $x \in X'$ and

$$f_1(X') = \sum_{j=1}^{k} \max\{|X_j \cap X'|, 1\}. \tag{2.3}$$

Suppose $\mathcal{F} \neq \emptyset$. Since $\mathcal{F}$ is closed under unions and intersections (as the left- and right-hand sides of (2.3) are submodular and supermodular, respectively), there is a unique maximal element Y in $\mathcal{F}$. If Y intersects each of the X_i then

$$f_1(Y) = \sum_{j=1}^{k} |X_j \cap Y| = |Y \cap (X_1 \cup \cdots \cup X_k)| \leq |Y \setminus \{x\}| < |Y|, \tag{2.4}$$

contradicting (2.1). So without loss of generality we may assume that $Y \cap X_1 = \emptyset$. This implies that,

$$\text{if } x \in X' \text{ and } X' \cap X_1 \neq \emptyset, \text{ then } f_1(X') > \sum_{j=1}^{k} \max\{|X_j \cap X'|, 1\}. \tag{2.5}$$

Obviously, this is also true if $\mathcal{F} = \emptyset$.

Similarly, there exists an index j such that

$$\text{if } x \in X' \text{ and } X' \cap X_j \neq \emptyset, \text{ then } f_2(X') > \sum_{j=1}^{k} \max\{|X_j \cap X'|, 1\}. \tag{2.6}$$

If $j = 1$ one easily checks that replacing X_1 by $X_1 \cup \{x\}$ does not violate (2.2), contradicting the maximality of $X_1 \cup \cdots \cup X_k$. So suppose $j \neq 1$, say $j = 2$.

Now (2.5) and (2.6) imply

$$f_i(X') \geq \max\{|(X_1 \cup X_2 \cup \{x\}) \cap X'|, 2\} + \sum_{j=3}^{k} \max\{|X_j \cap X'|, 1\} \tag{2.7}$$

for each nonempty subset X' of X, and $i = 1, 2$. Define

$$f_i'(X') = \min_{X'' \subset X \setminus (X_1 \cup X_2 \cup \{x\})} f_i(X' \cup X'') - \sum_{j=3}^{k} \max\{|X_j \cap X''|, 1\} \tag{2.8}$$

for subsets X' of $X_1 \cup X_2 \cup \{x\}$, and $i = 1, 2$. The functions f_1' and f_2' are submodular again, and from (2.7) we know that

$$f'_i(X') \geq \max\{|X'|, 2\} \tag{2.9}$$

for each nonempty subset X' of X, and $i = 1, 2$. Hence, by part (i) above, we can split $X_1 \cup X_2 \cup \{x\}$ into classes X'_1 and X'_2 such that

$$f'_i(X') \geq \sum_{j=1}^{2} \max\{|X'_j \cap X'|, 1\} \tag{2.10}$$

for each nonempty subset X' of $X_1 \cup X_2 \cup \{x\}$, and $i = 1, 2$. Hence, by definition (2.8) of the f'_i, the sets $X'_1, X'_2, X'_3 = X_3, \ldots, X'_k = X_k$ form a collection of pairwise disjoint sets satisfying

$$f_i(X') \geq \sum_{j=1}^{k} \max\{|X'_j \cap X'|, 1\} \tag{2.11}$$

for each nonempty subset X' of X, contradicting the maximality of $X_1 \cup \cdots \cup X_k$. $\square$

In fact, Theorem 2 may be considered as a theorem on the splitting of vectors in polymatroids (cf. [1]), since it can be extended easily to: *let f_1 and f_2 be integral submodular set-functions on a set X, and let $b : X \to \mathbb{Z}_+$ be such that $f_i(X') \geq \max\{b(X'), k\}$ for each nonempty subset X' of X, and $i = 1, 2$. Then there exist $b_1, \ldots, b_k : X \to \mathbb{Z}_+$ such that $b = b_1 + \cdots + b_k$ and $f_i(X') \geq \Sigma_{j=1}^{k} \max\{b_j(X'), 1\}$ for each nonempty subset X' of X, and $i = 1, 2$.*

3. A bi-branching theorem

Combination of Theorem 1 and Theorem 2 gives a theorem on bi-branchings.

Theorem 3. *Let $D = (V, A)$ be a directed graph, and let V be split into classes V_1 and V_2, such that any nonempty subset of V_1 (of V_2, respectively) is entered (left, respectively) by at least k arrows of D. Then A can be split into k bi-branchings.*

Proof. Let X be the set of arrows from V_2 to V_1, and define the set-functions f_1 and f_2 on X by

$$f_1(X') = \min\{d_A^-(V_1') \mid V_1' \subset V_1, \text{ and each arrow in } X' \text{ ends in } V_1'\},$$
$$f_2(X') = \min\{d_A^+(V_2') \mid V_2' \subset V_2, \text{ and each arrow in } X' \text{ starts in } V_2'\}, \tag{3.1}$$

for $X' \subset X$. It is easy to check that f_i is submodular, and that $f_i(X') \geq \max\{|X'|, k\}$, for each nonempty subset X' of X, and $i = 1, 2$. Hence, by Theorem 2, we can split X into classes $X_1, \ldots, X_k$ such that

$$f_i(X') \geq \sum_{j=1}^{k} \max\{|X_j \cap X'|, 1\} \tag{3.2}$$

for each nonempty subset X' of X, and $i = 1, 2$. Let Y_j (Z_j, respectively) be the set of heads (tails, respectively) of arrows occurring in X_j. Consider any nonempty subset V_1' of V_1, and let X_0 be the set of arrows in X with head in V_1'. If $X_0 \neq \emptyset$, by (3.2)

$$f_1(X_0) \geq \sum_{j=1}^{k} \max\{|X_j \cap X_0|, 1\}. \tag{3.3}$$

In particular,

$$d_A^-(V_1') \geq f_1(X_0) \geq |X_0| + |\{j | X_j \cap X_0 = \emptyset\}| = |X_0| + h(V_1'), \tag{3.4}$$

where $h(V_1')$ is the number of j with $Y_j \cap V_1' = \emptyset$. Hence, as $d_X^-(V_1') = |X_0|$, it follows that

$$d_{A'}^-(V_1') \geq h(V_1'), \tag{3.5}$$

where A' is the set of arrows contained in V_1. As (3.5) is true also if $X_0 = \emptyset$, (3.5) is true for each nonempty subset V_1' of V_1, and hence, by Theorem 1 we can split A' into classes $A_1', \ldots, A_k'$ such that if V_1' is a nonempty subset of $V_1 \backslash Y_j$, then at least one arrow in A_j' enters V_1', for $j = 1, \ldots, k$.

Similarly, one can split the arrows contained in V_2 into k classes $A_1'', \ldots, A_k''$ such that if V_2' is a nonempty subset of $V_2 \backslash Z_j$ then at least one arrow in A_j'' leaves V_2', for $j = 1, \ldots, k$.

It follows that $A_1' \cup X_1 \cup A_1'', \ldots, A_k' \cup X_k \cup A_k''$ yields a splitting as required. □

4. A special case of a conjecture of Edmonds and Giles

Theorem 3 is used to show the following theorem, which proves a special case of a conjecture of Edmonds and Giles [3], announced independently by D.H. Younger.

Theorem 4. *Let $D = (V, A)$ be an acyclic directed graph, such that any pair of source and sink is connected by a directed path, and let C be a subset of A such that each directed cut of D intersects C in at least k arrows. Then C can be split into classes $C_1, \ldots, C_k$ such that each class C_i intersects each directed cut.*

Proof. We prove the theorem by induction on $|V| + |C|$. Suppose the assertion does not hold for D and C, and suppose this counterexample has been chosen with $|V| + |C|$ as small as possible.

Call a subset V' of V a *kernel* for D if $\emptyset \neq V' \neq V$ and $d_A^+(V') = 0$. We may assume without loss of generality that if there is a directed path in D from v to w, then $(v, w) \in A$, as the adding of such arrows does not change the collection of kernels. So we may think of D as just a partially ordered set.

(i) *If V' is a kernel with $d_C^-(V') = k$, then $|V'| = 1$ or $|V \backslash V'| = 1$, i.e., directed cuts intersecting C in exactly k arrows are determined by sources and sinks.* For suppose V' is a kernel with $d_C^-(V') = k$ and $|V'| \geq 2$ and $|V \backslash V'| \geq 2$. Let C' be the set of arrows in C with head in V', and let C'' be the set of arrows in C with tail in $V \backslash V'$. Contracting $V \backslash V'$ to one point yields a smaller directed graph $D' = (V', A')$, with $C' \subset A'$ and each directed cut of D' intersecting C' in at least k arrows. Hence, by induction, C' can be split into classes $C'_1, \ldots, C'_k$ such that each directed cut of D' intersects each C'_i. Similarly, by contracting V', thus obtaining the directed graph D'', the projection C'' of C can be split into classes $C''_1, \ldots, C''_k$ such that each directed cut of D'' intersects each C''_i. So each C'_i and each C''_i contain exactly one of the arrows in C from $V \backslash V'$ to V', and we may assume that $C'_i \cap C''_j \neq \emptyset$ if and only if $i = j$. Therefore, the sets $C'_1 \cup C''_1, \ldots, C'_k \cup C''_k$ partition C, and for any kernel V'' of D with $V'' \subset V'$ or $V' \subset V''$ or $V' \cap V'' = \emptyset$ or $V' \cup V'' = V$ there is an arrow in $C'_i \cup C''_i$ entering V'', for each i. To prove that this is true for *each* kernel of D, let V'' be a kernel with

$$d^-_{C'_i \cup C''_i}(V'') = 0 \tag{4.1}$$

for a certain i. So $V' \cap V'' \neq \emptyset$ and $V' \cup V'' \neq V$, and hence $V' \cap V''$ and $V' \cup V''$ are kernels of D again. Also

$$d^-_{C'_i \cup C''_i}(V' \cap V'') + d^-_{C'_i \cup C''_i}(V' \cup V'') \leq d^-_{C'_i \cup C''_i}(V') + d^-_{C'_i \cup C''_i}(V''). \tag{4.2}$$

Since $d^-_{C'_i \cup C''_i}(V') = 1$, at least one of the two left terms is 0. But $V' \cap V'' \subset V' \subset V' \cup V''$, and hence both left terms are nonzero.

(ii) *If $a = (v, w)$ belongs to C, then v is a source of D or w is a sink of D.* For suppose not. Then, by (i), a is not in any directed cut intersecting C in exactly k arrows. So removing a from C, by induction, $C \backslash \{a\}$ can be split into k coverings for the directed cuts. Hence also C can be split in such a way.

(iii) *If* $a = (v, w)$ *and* $a' = (v', w')$ *belong to* C, *and* (v', w) *belongs to* A, *then* v' *is a source or* w *is a sink of* D. For suppose v' is not a source and w is not a sink. By (ii) this implies that v is a source and w' is a sink, and hence $a'' = (v, w')$ belongs to A. Since $a \neq a'$ (as v is a source and v' not), the set $C' = (C \setminus \{a, a'\}) \cup \{a''\}$ is smaller than C. Moreover, $d^-_{C'}(V') \geq k$ for each kernel V' of D, as the number of arrows in C' meeting any source or sink is the same as that for C, and, in general,

$$d^-_{\{a''\}}(V') \leq d^-_{\{a,a'\}}(V') \leq d^-_{\{a''\}}(V') + 1 . \tag{4.3}$$

So, by induction, C' can be split into classes $C'_1, \ldots, C'_k$ such that $d^-_{C'_i}(V') \geq 1$ for each kernel V' and each i. Assuming $a'' \in C'_1$, we can replace C'_1 by $(C'_1 \setminus \{a''\}) \cup \{a, a'\}$, and this yields, by (4.3), a splitting of C as required.

(iv) *There exists a kernel* V' *for* D, *containing all sinks but no sources, such that if* $(v, w) \in C$ *enters* V', *then* v *is a source and* w *is a sink*. For let V' consist of all sinks, together with all vertices u for which there is an arrow (v, w) in C with v not a source, and $(v, u) \in A$. One easily checks that V' is a kernel containing all sinks but no sources. Moreover, suppose $(t, u) \in C$ enters V'. If u is a sink and t is no source, then $t \in V'$, contradicting that (t, u) enters V'. If t is a source and u is not a sink, then, by definition of V', there is an arrow (v, w) in C with v not a source, and $(v, u) \in A$. But this contradicts (iii). Hence t is a source and u is a sink.

(v) Let V' be as in (iv), and let $V'' = V \setminus V'$. Let $D' = (V, A')$ be the directed graph arising from D by replacing any arrow (v, w) of D by k parallel arrows from w to v. One easily checks that

$$d^-_{A' \cup C}(W) \geq k \tag{4.4}$$

for each nonempty proper subset W of V. So, by Theorem 3, the set $A' \cup C$ can be split into k bi-branchings with respect to the splitting V', V''. Let $C_1, \ldots, C_k$ be the intersections of these bi-branchings with C. Hence

$$d^-_{A' \cup C_j}(W) \geq 1 \tag{4.5}$$

for each nonempty proper subset W of V with $W \subset V'$ or $V' \subset W$, and $j = 1, \ldots, k$. We show that each C_j intersects each directed cut, which finishes our proof.

Let W be a kernel for D, and let $j = 1, \ldots, k$. We prove that at least one arrow in C_j enters W. Note that if W contains any source, it contains all sinks.

First suppose that W contains no sources of D. By (4.5),

$$d^-_{A' \cup C_j}(W \cap V') \geq 1 . \tag{4.6}$$

Since $W \cap V'$ again is a kernel of D, we have $d^-_{A'}(W \cap V') = 0$, and hence there is an arrow in C_j entering $W \cap V'$. Since, by (iv), each arrow in C entering V' starts in a source, and since W does not contain any source, this arrow enters W also.

Second, if W contains every sink of D, by symmetric arguments (now considering $W \cup V'$) again at least one arrow of C_j enters W. □

Remark. There is another special case in which the conjecture of Edmonds and Giles is true, namely if D arises from a directed tree T, with vertex set V, by taking the transitive closure (i.e., $A = \{(v, w) \mid$ there is a directed path in T from v to $w\}$. This can be shown using the total unimodularity of matrices involved. One may ask for a common generalization of this special case and Theorem 4 above.

5. An extension of Theorem 4

We now extend Theorem 4, thus obtaining a common generalization of the Theorems 3 and 4 (cf. Section 0), by the following observation.

Observation. Let $D = (V, A)$ and $D' = (V, A')$ be directed graphs. Let $a = (v, w)$ be an arrow of D such that there exist vertices v' and w', and directed paths in D' from v to v', from w' to v', and from w' to w (cf. Fig. 3, where wriggled lines represent directed paths in D'). The vertices v, v', w', w need not to be distinct.

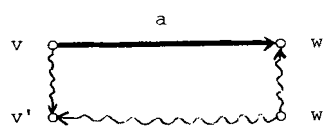


Fig. 3.

Now let v'' and w'' be two new vertices, let $V_0 = V \cup \{v'', w''\}$, $a'' = (v'', w'')$, $A_0 = (A \backslash \{a\}) \cup \{a''\}$, and $A'_0 = A' \cup \{(v, v''), (v'', v'), (w'', v''), (w', w''), (w'', w)\}$, $D_0 = (V_0, A_0)$, $D'_0 = (V_0, A'_0)$ (cf. Fig. 4, where heavy and light lines stand for arrows of D_0 and D'_0, respectively). Then one easily checks that, for each subset A'' of A, A'' is a strong connector for D', if and only if A''_0 is a strong connector for D'_0, where $A''_0 = A''$ if $a \notin A''$, and $A''_0 = (A'' \backslash \{a\}) \cup \{a''\}$ if $a \in A''$. Hence the hypergraphs of strong connectors for the two cases are isomorphic. Therefore, also the hypergraphs of minimal strong cuts are isomorphic (as these are the 'blockers' of the first ones).

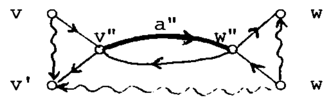


Fig. 4.

This gives us the invariance of certain min–max relations under these transformations. We shall apply this observation to derive the following theorem from Theorem 4, and Theorem 6 from the Lucchesi–Younger theorem.

Theorem 5. *Let $D' = (V, A')$ be an acyclic directed graph, such that each pair of source and sink is connected by a directed path. Let $D = (V, A)$ be a directed graph. Then the maximum number of pairwise disjoint strong connectors for D' is equal to the minimum size of a strong cut induced by D'.*

Proof. We may suppose that D' is transitive, i.e., that if (u, v) and (v, w) are in A', then also (u, w) is in A'. We prove the theorem by induction on the number of arrows $a = (v, w)$ in A with (w, v) not in A'. If this number is 0, the theorem is equivalent to Theorem 4.

So suppose $a = (v, w) \in A$ and $(w, v) \notin A'$. Let v' be a sink of D' with $(v, v') \in A'$, and let w' be a source of D' with $(w', w) \in A'$. By assumption $(w', v') \in A'$, and hence we may make digraphs D_0 and D_0' as in the Observation above. Since strong connectors, and strong cuts, determine isomorphic hypergraphs in the two cases, the conditions of the theorem hold also for D_0 and D_0'. Since the number of arrows in D_0 which do not occur in reversed direction in D_0', is one less than for D and D', we can split A_0 as required, and hence, since the hypergraphs are isomorphic, we can split A as required. □

One easily derives the following weighted version.

Corollary 5a. *Let $D' = (V, A')$ be an acyclic directed graph, such that each pair of source and sink is connected by a directed path. Let $D = (V, A)$ be a directed graph, and let $c : A \to \mathbb{Z}_+$ be a capacity function. Suppose that the minimum capacity of a strong cut induced by D' is at least k. Then there are k strong connectors for D' such that no arrow a is in more than $c(a)$ of these strong connectors.*

Proof. Replace each arrow a of D by $c(a)$ parallel arrows, and apply Theorem 5. □

6. A similar extension of the Lucchesi–Younger theorem

We can apply the Observation of Section 5 also to obtain a somewhat more general form of the Lucchesi–Younger theorem [18] (cf. [16]). The Lucchesi–Younger theorem says that the minimum size of a set of arrows in a directed graph $D = (V, A)$ intersecting each directed cut, is equal to the maximum number of pairwise disjoint directed cuts. It is easy to derive, by replacing arrows by directed paths, from this a weighted version: given a length function $l : A \to \mathbb{Z}_+$, the minimum length of a set of arrows intersecting all directed cuts, is equal to the maximum number of directed cuts such that no arrow a is in more than $l(a)$ of these directed cuts.

The more general theorem is as follows.

Theorem 6. *Let $D = (V, A)$ and $D' = (V, A')$ be directed graphs, such that for each arrow $a = (v, w)$ of D there are vertices v' and w' and directed paths in D' from v to v', from w' to v', and from w' to w. Let $l : A \to \mathbb{Z}_+$ be a length function. Then the minimum length of a strong connector for D' is equal to the maximum number of strong cuts induced by D' such that no arrow a is in more than $l(a)$ of these strong cuts.*

Proof. The proof is similar to that of Theorem 5. □

A direct corollary is another 'bi-branching theorem'.

Corollary 6a. *Let $D = (V, A)$ be a directed graph, and let V be split into classes V' and V''. Let $l : A \to \mathbb{Z}_+$ be a length function such that each bi-branching has length at least k. Then there are nonempty proper subsets $V_1, \ldots, V_k$ of V such that $V_i \subset V'$ or $V' \subset V_i$ for each i, and no arrow a enters more than $l(a)$ of the V_i.*

Proof. Apply Theorem 6, with $A' = \{(v', v'') \mid v' \in V', v'' \in V''\}$. □

Direct consequences to Corollary 6a are Fulkerson's branching theorem [9] and König's theorem [15] on minimal coverings in bipartite graphs. Note that, conversely, the cardinality version of Corollary 6a (i.e., $l \equiv 1$) can be derived easily from König's theorem.

7. Sub- and supermodular functions on directed graphs

Edmonds and Giles [3] gave a common generalization of the Lucchesi–Younger theorem [18] (cf. Section 6) and Edmonds' matroid intersection theorem [1], by considering submodular functions defined on the vertex set of a directed graph. In fact, also the extension of the Lucchesi–Younger theorem given above (Theorem 6) may be included in such a framework—see Theorem 7 below.

Note that a collection $\mathcal{F}$ of subsets of a set V, containing $\emptyset$ and V, is closed under unions and intersections, if and only if there is a directed graph $D' = (V, A')$ such that $\mathcal{F} = \{V' \subset V \mid d^-_{A'}(V') = 0\}$. The following theorem extends Theorem 6 above and another theorem of Frank [5].

Theorem 7. *Let $\mathcal{F}$ be a collection of subsets of V and let f be an integral function defined on $\mathcal{F}$, such that if $V_1, V_2 \in \mathcal{F}$ and $V_1 \cap V_2 \neq \emptyset$, $V_1 \cup V_2 \neq V$, then $V_1 \cap V_2 \in \mathcal{F}$, $V_1 \cup V_2 \in \mathcal{F}$ and $f(V_1 \cap V_2) + f(V_1 \cup V_2) \geq f(V_1) + f(V_2)$. Let furthermore a directed graph $D = (V, A)$ be given such that if $V_1, V_2, V_3 \in \mathcal{F}$ with $V_1 \cap V_2 \cap V_3 = \emptyset$ and $V_1 \cup V_2 \cup V_3 = V$, then no arrow of D enters both V_1 and V_2. Let $l : A \to \mathbb{Z}_+$ be a length function. Then the minimum length of a set $A'' \subset A$ such that each $V' \in \mathcal{F}\backslash\{\emptyset, V\}$ is entered by at least $f(V')$ arrows in A'', is equal to the maximum value of*

$$\sum_{i=1}^{k} f(V_i), \tag{7.1}$$

where $V_1, \ldots, V_k$ are sets in $\mathcal{F}\backslash\{\emptyset, V\}$ such that each arrow a of D enters at most $l(a)$ of the V_i.

(The theorem asserts that both sides of a certain linear programming duality equation are achieved by integral solutions—cf. Section 8.)

Theorem 7 can be proved with the standard methods (using cross-free collections, tree-representations, total dual integrality), as described by Edmonds and Giles [3].

Note that the condition given in the second sentence of Theorem 7 is just the analogue of the condition given in the first sentence of Theorem 6. In order to obtain a similar generalization of Theorem 5, one easily checks that a collection $\mathcal{F}$, closed under unions and intersections, is the collection of sets V' with $d^-_{A'}(V') = 0$ for some digraph $D' = (V, A')$ with the property that, after contracting the strong components of D', each pair of source and sink is connected by a directed path, if and only if there are no sets V_1, V_2, V_3 in $\mathcal{F}\backslash\{\emptyset, V\}$ with $V_1 \cap V_2 \cap V_3 = \emptyset$ and $V_1 \cup V_2 \cup V_3 = V$.

Now the following possible generalization of Theorem 5 is *not true*: let $\mathcal{F}$ be a collection of subsets of V with the properties described in the previous paragraph, let f be a supermodular function on $\mathcal{F}$, and let $D = (V, A)$ be a directed graph, such that each set V' in $\mathcal{F}\backslash\{\emptyset, V\}$ is entered by at least $f(V')$

arrows of D. Suppose $f=f_1+f_2$, where f_1 and f_2 are nonnegative integral supermodular functions on $\mathcal{F}$. Then A can be split into classes A_1 and A_2 such that each set V' in $\mathcal{F}\backslash\{\emptyset, V\}$ is entered by at least $f_i(V')$ arrows in A_i, for $i=1,2$. A counterexample to this is given by taking D as in Fig. 5, $\mathcal{F}$ being the collection of all subsets of $V\backslash\{r\}$, and $f=f_1+f_2$, where, for V' in $\mathcal{F}$, $f_1(V')=1$, $f_2(V')=1$ if $s\in V'$, and $f_2(V')=0$ if $s\notin V'$.

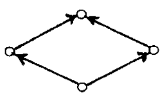


Fig. 5.

The following generalization of Theorem 2 and Theorem 5 might be true.

> Let $\mathcal{F}$ be a collection of subsets of a set V, closed under unions and intersections, such that for no V_1, V_2, V_3 in $\mathcal{F}\backslash\{\emptyset, V\}$ both $V_1\cap V_2\cap V_3=\emptyset$ and $V_1\cup V_2\cup V_3=V$. Let f be a submodular function on $\mathcal{F}$ such that $f(V')\geq k$ for each V' in $\mathcal{F}\backslash\{\emptyset, V\}$. Let $D=(V,A)$ be a directed graph such that $d_A^-(V')\leq f(V')$ for each V' in $\mathcal{F}\backslash\{\emptyset, V'\}$. Then A can be split into classes $A_1,\ldots,A_k$ such that for each V' in $\mathcal{F}\backslash\{\emptyset, V\}$ one has $\Sigma_{j=1}^{k}\max\{d_{A_j}^-(V'),1\}\leq f(V')$. (7.2)

By taking $f(V')=d_A^-(V')$ Theorem 5 follows. By taking A to be a collection of disjoint arrows, with set V' of heads, and $\mathcal{F}$ to be the collection of all V'' with $V''\subset V'$ or $V'\subset V''$, Theorem 2 follows.

The question remains whether both the generalizations of Edmonds–Giles type, and assertions of the type of Theorem 2 and problem (7.2) above, fit into one framework. Also at another point submodular functions, or rather matroids, appear, namely at Fulkerson's branching theorem. This theorem may be interpreted as a min–max relation for the minimum weight of a common base of two matroids (cf. [1]). One may ask whether the more general bi-branching theorem (Corollary 6a), or even Theorem 6, can be formulated in such a way.

8. Polyhedral representations and polynomial algorithms

As usual with min–max relations, Theorems 5 and 6 above allow a polyhedral formulation, or, equivalently, a formulation in terms of linear programming. By the ellipsoid method as described in [10] this often yields the existence of polynomial algorithms.

Let D and D' be as in Theorem 6, and let $c : A \to \mathbb{Z}_+$. Consider the linear programming problem of finding

$$\min \sum_{a \in A} c(a)x(a) \tag{8.1}$$

where $x : A \to \mathbb{Q}_+$ such that

$$\begin{aligned} &0 \leq x(a) \leq 1 \quad \text{if } a \in A\,, \\ &\sum_{a \in A''} x(a) \geq 1 \quad \text{if } A'' \text{ is a strong cut}\,, \end{aligned} \tag{8.2}$$

where we mean a 'strong cut' to be induced by D'. By the Duality Theorem of linear programming, (8.1) is equal to

$$\max \sum_{A'' \text{ strong cut}} y(A'')\,, \tag{8.3}$$

where, for each strong cut A'', $y(A'')$ is a rational number such that

$$\begin{aligned} &y(A'') \geq 0\,, \quad \text{if } A'' \text{ is a strong cut,} \\ &\sum_{A'' \ni a} y(A'') \leq c(a)\,, \quad \text{if } a \in A. \end{aligned} \tag{8.4}$$

Now, Theorem 6 asserts that (8.1) and (8.3) are attained by *integral* functions x and y. So the system of linear inequalities (8.2) is totally dual integral (cf. [3]), and a function x satisfies (8.2) if and only if x is a convex linear combination of incidence vectors of strong connectors for D'.

If, moreover, D' is acyclic and each pair of source and sink of D' is connected by a directed path, we obtain similar conclusions if we exchange the terms 'strong cut' and 'strong connector', as follows from Corollary 5a. Note that in the latter case, by the theory of blocking polyhedra of Fulkerson [8], if D and D' satisfy the weaker conditions of Theorem 6 only, (8.1) is attained by an integral vector x (i.e., by the incidence vector of some strong cut).

Therefore, by the ellipsoid method there exists a polynomial algorithm for finding minimum length strong connectors, if and only if there exists a polynomial algorithm for finding minimum capacitated strong cuts. However, the existence of the latter algorithm follows easily from the Ford–Fulkerson min-cut algorithm (by giving the arrows of D' sufficiently large capacity), and hence minimum length strong connectors can be found in polynomial time. Also a maximum packing of strong cuts (i.e., an integer solution for (8.3)) can be found in polynomial time, by applying the usual techniques of making cuts cross-free (cf. [10]). Clearly, minimum length strong connectors and maximum packings of strong cuts can be found also by adapting (e.g., by the Observation of Section 5) the existing polynomial algorithms for the Lucchesi–Younger theorem [6, 12, 17].

It remains to show that the splitting of A as described in Theorem 5 and Corollary 5a can be found efficiently. However, our proof above yields a polynomial algorithm. Indeed, the proof of Theorem 5 reduces this theorem to Theorem 3. Since this reduction can be carried out in polynomial time, we need to show that a splitting into bi-branchings can be found efficiently. But the splitting into bi-branchings is obtained by first splitting the 'crossing arrows' (from V_2 to V_1), which splitting can be found by Theorem 2. After that this splitting is extended to a splitting into bi-branchings by Theorem 1. Now to derive polynomial algorithms from the proofs of Theorem 1 and Theorem 2, one needs only a method to find one, or all, minimal nonempty subsets V' with $f(V') = h(V')$, where f is submodular and h is supermodular, with $h \leq f$. But this can be reduced easily to the problem of finding a set minimizing a submodular set-function, and this can be solved in polynomial time [10].

Also the splitting described in Corollary 5a, i.e., an integral solution y for (8.3), with strong connectors instead of strong cuts, can be found in time polynomially bounded by the size of the problem. Note that this size is

$$|V| + |A| + |A'| + \sum_{a \in A} \log(c(a) + 1), \tag{8.5}$$

so, to obtain a good algorithm, we cannot just replace each arrow a by $c(a)$ parallel arrows. However, by the ellipsoid method a fractional solution y of (8.3) (again with 'strong connector' instead of 'strong cut'), can be obtained in polynomial time, such that the number of strong connectors A'' with $y(A'') > 0$ is at most $|A|$. Now let

$$c'(a) := \sum_{A'' \ni a} (y(A'') - \lfloor y(A'') \rfloor), \tag{8.6}$$

where the sum ranges over strong connectors A'', and where $\lfloor\ \rfloor$ denotes lower integer part. Since $c'(a) \leq |A|$ we can replace each arrow a by $c'(a)$ parallel arrows, and then find in this new directed graph as many as possible pairwise disjoint strong connectors, by the method described above for Theorem 5, i.e., we find integers $y'(A'') \geq 0$ for each strong connector A''. One easily checks that $\lfloor y(A'') \rfloor + y'(A'')$ is an integer solution for (8.3).

Acknowledgement

We thank Dr. A. Frank (Budapest) and Prof. L. Lovász (Szeged) for their helpful comments.

References

[1] J. Edmonds, Submodular functions, matroids, and certain polyhedra, in: R. Guy, et al., eds., Combinatorial Structures and their Applications (Gordon and Breach, New York, 1970) pp. 69–87.
[2] J. Edmonds, Edge-disjoint branchings, in: B. Rustin, ed., Combinatorial Algorithms (Academic Press, New York, 1973) pp. 91–96.
[3] J. Edmonds and R. Giles, A min–max relation for submodular functions on graphs, in: P.L. Hammer, et al., eds., Studies in Integer Programming, Ann. of Discrete Math. 1 (1977) 185–204.
[4] E. Egerváry, Matrixok kombinatorius tulajdonságairol, Mat. Fiz. Lapok 38 (1931) 16–28.
[5] A. Frank, Kernel systems of directed graphs, Acta Sci. Math. (Szeged) 41 (1979) 63–76.
[6] A. Frank, How to make digraph strongly connected, Combinatorica 1 (1981) 145–153.
[7] D.R. Fulkerson, Networks, frames, and blocking systems, in: G.B. Dantzig and A.F. Veinott, eds., Mathematics of the Decision Sciences, Part I (Amer. Math. Soc., Providence, R.I., 1968) pp. 303–334.
[8] D.R. Fulkerson, Blocking polyhedra, in: B. Harris, ed., Graph Theory and its Applications (Academic Press, New York, 1970) pp. 93–112.
[9] D.R. Fulkerson, Packing rooted cuts in a weighted directed graph, Math. Programming 6 (1974) 1–13.
[10] M. Grötschel, L. Lovász and A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, Combinatorica 1 (1981) 169–197.
[11] R.P. Gupta, A decomposition theorem for bipartite graphs, in: P. Rosenstiehl, ed., Theory of Graphs (Gordon and Breach, New York, 1967) pp. 135–138.
[12] A.V. Karzanov, On the minimal number of arcs of a digraph meeting all its directed cutsets, to appear.
[13] D. König, Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre, Math. Ann. 77 (1916) 453–465.
[14] D. König, Graphok és matrixok, Mat. Fiz. Lapok 38 (1931) 116–119.
[15] D. König, Ueber trennende Knotenpunkte in Graphen (nebst Anwendungen auf Determinanten und Matrizen), Acta Lit. Sci. Sect. Sci. Math. (Szeged) 6 (1932–1934) 155–179.
[16] L. Lovász, On two minimax theorems in graph theory, J. Combin. Theory Ser. B 21 (1976) 96–103.
[17] C.L. Lucchesi, A minimax equality for directed graphs, D. Th. University of Waterloo, Waterloo, Ont., 1976.
[18] C.L. Lucchesi and D.H. Younger, A minimax relation for directed graphs, J. London Math. Soc. (2)17 (1978) 369–374.
[19] K. Menger, Zur allgemeinen Kurventheorie, Fund. Math. 10 (1927) 96–115.
[20] A. Schrijver, A counterexample to a conjecture of Edmonds and Giles, Discrete Math. 32 (1980) 213–214.
[21] D. de Werra, Some remarks on good colorations, J. Combin. Theory Ser. B 21 (1976) 57–64.

Annals of Discrete Mathematics 16 (1982) 281–286

THE BANDWIDTH PROBLEM: CRITICAL SUBGRAPHS AND THE SOLUTION FOR CATERPILLARS

Maciej M. SYSŁO and Jerzy ŻAK

Institute of Computer Science, University of Wrocław, 50 384 Wrocław, Poland

This paper contains the solution of the bandwidth problem for caterpillars. The method applied makes use of some lower bounds to the bandwidth of a graph in terms of subgraphs and their sizes and diameters. We introduce also several types of critical subgraphs related to the bandwidth problem and describe some of their properties.

1. Introduction

Suppose G is a graph without loops and multiple edges with vertex set $V(G)$ and edge set $E(G)$. Let $n(G)$ denote the number of vertices of a graph G. Although some results of this paper can be easily generalized for some infinite denumerable graphs, all graphs are assumed to be finite.

A *labeling* π is a one-to-one mapping from $V(G)$ to the positive integers N. The *bandwidth of* π is defined to be

$$B_\pi(G) = \max\{|\pi(u) - \pi(v)| : \{u, v\} \in E(G)\}.$$

We shall call $\pi(u)$ a *vertex label* and $|\pi(u) - \pi(v)|$ for $\{u, v\} \in E(G)$, an *edge label.*

The *bandwidth of a graph* G is defined to be

$$B(G) = \min\{B_\pi(G) : \pi \text{ is a labeling of } G\}.$$

Papadimitriou [4] proved that the problem of determining the bandwidth of a graph is NP-complete, and Garey et al. [3] showed that it remains NP-complete even when restricted to trees with maximum degree 3.

In Section 2 we discuss some new notions related to some subgraphs of a graph which play an important role in determining the bandwidth and, in Section 3, the bandwidth of an arbitrary caterpillar is evaluated.

The reader interested in other results on the bandwidth problem is referred to a survey paper [2].

2. Critical subgraphs

It is obvious that $B(G) \geq B(G')$ for every subgraph G' of G. Moreover, for many graphs G there exists a proper subgraph G' of G such that $B(G) = B(G')$, therefore some edges of G are immaterial for the value of the bandwidth of G. A subgraph G_B of G is said to be a *bandwidth-critical subgraph of* G if $B(G) = B(G_B)$ and $B(G_B) > B(G')$ for every proper subgraph G' of G_B.

Observation 2.1. *Every graph G has a bandwidth-critical subgraph.*

Proof. Let G' denote a subgraph of G with the minimum number of edges such that $B(G) = B(G')$, that is $B(G) > B(G' - e)$ for every edge $e \in E(G')$. Hence, $B(G') > B(G'')$ for every proper subgraph G'' of G'. Therefore G' is a bandwidth-critical subgraph of G. □

In fact, we proved that if G' is a bandwidth-critical subgraph of G, then $B(G') > B(G' - e)$ for every edge e of G'; such a graph G' is called *B-critical.* Therefore, a bandwidth-critical subgraph of a graph is B-critical.

It is unlikely that finding the bandwidth of a B-critical graph is an easier problem than that of finding the bandwidth of any graph and a characterization of such graphs seems to be also a difficult problem.

There are many lower bounds for the bandwidth of a graph (see [2]). Let $A(G)$ denote a lower bound to $B(G)$. Evidently we have

$$B(G) \geq \max\{A(G'): G' \text{ is a subgraph of } G\}.$$

Similarly to B-critical graphs we may also define graphs which are critical with respect to a lower bound. A graph G is *A-critical* if $A(G) > A(G - e)$ for every edge $e \in E(G)$. We have the following simple relation between B- and A-critical graphs.

Observation 2.2. *Let $B(G) = A(G')$, where A is a lower bound to B and $G' \subseteq G$. If G is B-critical, then $G' = G$ and G is also A-critical.*

This observation says that every B-critical graph for which there exists a lower bound A that determines its bandwidth is also A-critical.

Let us focus now our attention on the following lower bound for the bandwidth introduced by Chvátal (see [2]), which appears to be very useful in determining the bandwidth of many classes of trees (see Section 3)

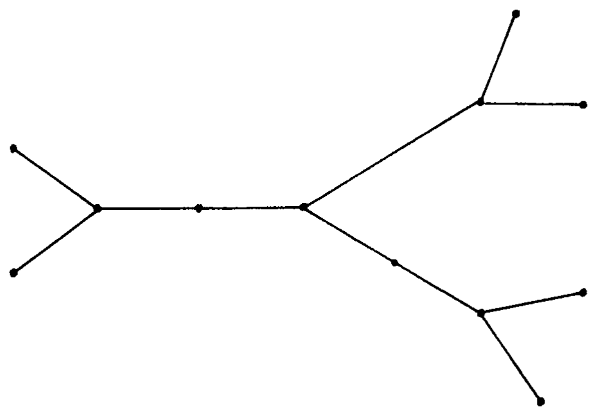
$$b(G)=\left\lceil\frac{n(G)-1}{d(G)}\right\rceil, \tag{1}$$

where $d(G)$ is the diameter of G and $\lceil a\rceil$ is the smallest integer not smaller than a.

Let for a given graph G, G_b denote a subgraph of G with the minimum number of edges and such that $b(G_b)=\max\{b(G')$: G' is a subgraph of $G\}$. Evidently, G_b is *b-critical*, that is $b(G_b)>b(G_b-e)$ for every edge $e\in E(G_b)$.

Fig. 1 shows a graph G for which $B(G)=3$ (why ?) and $B(G')\leqslant 2$ for every subgraph G' of G. It is easy to check that G is B-critical and G_b is isomorphic to $K_{1,3}$, therefore $B(G)>b(G')$ for every subgraph G' of G. It is an easy task to generalize the graph G to an infinite family of graphs $\{H_i\}_{i\geqslant 2}$ such that $H_2=G$, $b(H_i)=i$ and $B(H_i)\geqslant i+1$ for $i\geqslant 2$.

It is interesting that all b-critical subgraphs of connected graphs are trees.



G

Fig. 1.

Observation 2.3. *A b-critical subgraph of a connected graph is a tree.*

Proof. Let G be a connected graph and G' denote its b-critical subgraph. Evidently, G' is connected. Let p be a path in G' of length $d(G')$. The path p can be augmented to a spanning tree T of G' for which obviously $b(T)=b(G')$. Therefore, if G' is not a tree, it is not b-critical. $\square$

The last observation provides another motivation for our interest in the class of trees, which are the graphs that produce the best lower bound of the form (1). By this observation, one can write

$$B(G)\geqslant\max\{b(T)\colon T\subseteq G \text{ and } T \text{ is a tree}\}.$$

This relation between the bandwidth of G and its bound b can also be read as a weakness of b for arbitrary graphs.

Properties of critical graphs arising in the study of the bandwidth problem will be discussed in another paper (see [6] and also [2]). In Section 3 we present a new class of trees for which the bound b determines their bandwidth and gives rise to an efficient algorithm for its finding.

3. The bandwidth of caterpillars

The purpose of this section is to add one more subclass of trees to those for which the bandwidth problem can be easily solved. In [5] we solved the bandwidth problem for ordered caterpillars and in the sequel we show how to calculate the bandwidth of any caterpillar. Recently, we improved this result by providing a polynomial time algorithm for finding the bandwidth of a caterpillar with hairs of length 1 and 2 (see [1]).

A tree T is a *caterpillar* if removing all pendant vertices of T results in a path $p(T)$. Notice that $p(T)$ consists exactly of non-pendant vertices of T. A caterpillar T is *ordered* if the vertices of $p(T)$ are ordered by non-increasing degree. Assume that $p(T) = (w_1, w_2, \ldots, w_m)$ and let T_{ij} $(i \leq j)$ denote the subgraph of T consisting of $w_i, w_{i+1}, \ldots, w_{j-1}, w_j$ and of all their neighbours.

We proved in [5] that if T is an ordered caterpillar, then $B(T) = \max_{1 \leq i \leq m} b(T_{1,i})$ and the optimal labeling of T can be found in $O(n(T))$ time.

A similar result will now be proved for an arbitrary caterpillar.

Theorem 3.1. *If T is a caterpillar with m non-pendant vertices, then*

$$B(T) = \max_{1 \leq i,j \leq m} b(T_{ij}) .$$

Proof. Let E_i denote the set of pendant vertices adjacent to w_i, T_i be the subgraph induced by w_i and its neighbours, and let $b = \max_{1 \leq i,j \leq m} b(T_{ij})$.

We shall prove that for every caterpillar the following algorithm produces a labeling π of bandwidth b, which will complete the proof of the theorem.

Algorithm 3.2

Step 1. Let w_0 be one of the vertices in E_1. Set $\pi(w_0) = 1$.

Step 2. For $i = 1, 2, \ldots, m$ perform:

$$\text{(i) Set } \pi(w_i) = \min \begin{cases} \pi(w_{i-1}) + b, & \text{(i1)} \\ \max\limits_{v \in V(T_{i-1})} \pi(v) + |E_i| + 1 . & \text{(i2)} \end{cases}$$

(ii) Label the vertices of E_i using first $|E_i|$ unused integers from interval $[1, n(T)]$.

We show now that π is a labeling of $V(T)$ and satisfies $|\pi(u) - \pi(v)| \leqslant b$ for every $\{u, v\} \in E(T)$.

(a) Every label from interval $[1, n(T)]$ is used exactly once. To prove this fact, we show that $\pi(w_i) > \pi(u)$ for every $u \in V(T_{i-1}) - \{w_i\}$, $i = 1, 2, \ldots, m$. This is obvious if $\pi(w_i)$ is determined by (i2). If $\pi(w_i) = \pi(w_{i-1}) + b$, then it follows from the inequality $\pi(v) - \pi(w_{i-1}) < b$ for every $v \in E_{i-1}$ which is proved in (c).

(b) It follows from (i) that

$$0 < \pi(w_i) - \pi(w_{i-1}) \leqslant b \quad \text{for } i = 1, 2, \ldots, m.$$

(c) Since the vertices are labeled in the stars ordered as $T_1, T_2, \ldots, T_m$, it remains to prove that for every $i = 1, 2, \ldots, m-1$

$$|\pi(v) - \pi(w_{i+1})| < b$$

for every $v \in E_{i+1}$, that is $-b < \pi(v) - \pi(w_{i+1}) < b$.

To show the former inequality, note that, by Algorithm 3.2, $\pi(v) > \pi(w_i)$ for every $v \in E_{i+1}$. Therefore, by (b), we have

$$\pi(v) - \pi(w_{i+1}) > \pi(w_i) - \pi(w_{i+1}) \geqslant -b\,.$$

To show the latter inequality, assume that there exist $v \in E_{i+1}$ such that $\pi(v) - \pi(w_{i+1}) \geqslant b$. By Algorithm 3.2 we have $\pi(w_{i+1}) = \pi(w_i) + b$ since if $\pi(w_{i+1})$ is determined by (i2), then $\pi(v) < \pi(w_{i+1})$ for every $v \in E_{i+1}$. Let j be the smallest index such that $\pi(w_{l+1}) - \pi(w_l) = b$ holds for $l = j, j+1, \ldots, i-1, i$. Therefore, $\pi(w_j)$ is determined by (i2) and hence the number of vertices in $T_{1,j}$ is equal to $\pi(w_j) + 1$.

Assume that $\pi(w_j) = k + 1$, then $\pi(w_{j+1}) = k + 1 + b, \ldots, \pi(w_{i+1}) = k + 1 + (j - 1 + 1)b$.

By the theorem assumption we have

$$n(T_{j+1,i+1}) \leqslant b\, d(T_{j+1,i+1}) + 1 = b(j - i + 2) + 1\,.$$

Hence, the number of all vertices in $T_{1,i+1}$ is

$$\begin{aligned} n(T_{1,i+1}) &= n(T_{1,j}) + n(T_{j+1,i+1}) - 2 \\ &\leqslant k + 2 + b(j - i + 2) + 1 - 2 = k + b(j - i + 2) + 1\,. \end{aligned}$$

Since the maximum label for vertices in E_{i+1} is at most $n(T_{1,i+1})-1$ we obtain

$$\pi(v)-\pi(w_{i+1})\leqslant k+b(j-i+2)-(k+1+(j-i+1)b)=b-1<b,$$

what leads us to a contradiction with the assumption that there exist $v\in E_{i+1}$ such that $\pi(v)-\pi(w_{i+1})\geqslant b$.

Therefore, we showed that the labeling π of T is of bandwidth b. Since $b\leqslant B(T)$, we obtain $B(T)=b$. □

Corollary 3.3. *There is an* $O(n^2)$ *algorithm for finding the bandwidth of a caterpillar* T *with* n *vertices, which also produces a labeling achieving this bandwidth.*

Proof. Clearly, the bound b can be calculated in $O(n^2)$ time and the algorithm works in time proportional to n, so the total time used is $O(n^2)$. □

Example 3.4. For a caterpillar shown in Fig. 2 we have $B(T)=b(T_{1,5})=4$. The labeling indicated in Fig. 2 has been obtained by Algorithm 3.2 and is therefore optimal.

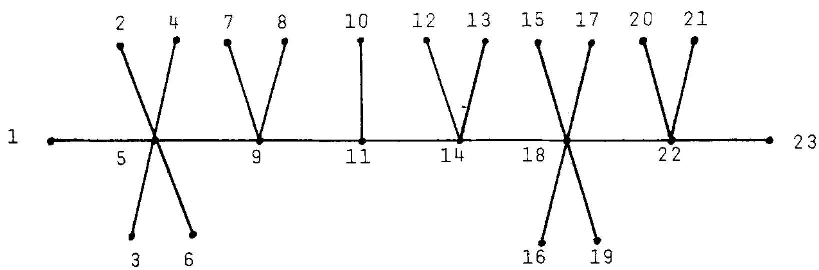


Fig. 2.

References

[1] S.F. Assmann, G.W. Peck, M.M. Sysło and J. Żak, The bandwidth of caterpillars with hairs of length 1 and 2, SIAM J. Algebraic and Discrete Methods 2 (1981) 387–393.

[2] P.Z. Chinn, J. Chvátalová, A.K. Dewdney and N.E. Gibbs, Graph bandwidth: a survey of theory and applications, J. Graph Theory, submitted.

[3] M.R. Garey, R.L. Graham, D.S. Johnson and D.E. Knuth, Complexity results for bandwidth minimization, SIAM J. Appl. Math. 34 (1978) 477–495.

[4] C.H. Papadimitriou, The NP-completeness of the bandwidth minimization problem, Computing 16 (1976) 263–270.

[5] M.M. Sysło and J. Żak, The bandwidth problem for ordered caterpillars, Tech. Rept. No. CS-80-065, Washington State University, 1980.

[6] J. Żak, The bandwidth problem for trees, Master Thesis, University of Wrocław, 1980.

Annals of Discrete Mathematics 16 (1982) 287–309

MINIMIZATION OF SOME NONLINEAR FUNCTIONS OVER POLYMATROIDAL NETWORK FLOWS

U. ZIMMERMANN

Mathematisches Institut, Universität zu Köln, Weyertal 86–90, D-5 Köln 41, W. Germany

We show the equivalence of some combinatorial models which generalize group-valued polymatroid intersections as well as group-valued network flows. We develop a negative circuit method for the minimization of certain nonlinear functions on such combinatorial structures. The method is an application of a general strategy proposed for minimizing certain nonlinear functions on a subset of R^n where R is a totally ordered ring.

1. Introduction

We consider the minimization of certain functions $f: D \to T$ where $D \subseteq R^n$, R is a totally ordered ring and T is a totally ordered set. We denote the set of feasible solutions by P. For solving

$$\min\{f(x) \mid x \in P\} \tag{1.1}$$

we develop a strategy which basically coincides with the primal simplex method of linear programming if we assume that $R = \mathbb{R}$, that $f: \mathbb{R}^n \to \mathbb{R}$ is linear, and that P is a polytope. On the other hand, our aim is to apply the strategy when the feasible solutions are generalizations of polymatroid intersections and network flows.

For real-valued linear objective functions ($f: \mathbb{R}^n \to \mathbb{R}$) such problems are discussed in several papers. Edmonds [1] introduced the polymatroid intersection problem; its capacitated version and further generalizations are discussed in [2, 8]. Martel [16] and Lawler [14] develop polymatroidal flows generalizing polymatroid intersections and network flows. They mention a thesis of Hassin [10] on the same problem. Recently, Lawler and Martel [15] developed an augmenting path method for the determination of maximal polymatroidal flows which needs no more than $O(n^3)$ augmentations ($R = \mathbb{R}$). Fujishige [7] describes independent flows which also generalize polymatroid intersections as well as network flows. That paper is related to previous work in [5, 6, 11, 12]. Fujishige [7] proposes two methods for solving the minimum cost–maximum independent flow problem ($f: R^n \to R$ linear in a totally ordered ring R): an augmenting path method and a primal negative circuit method. These methods are finite if $R = \mathbb{Z}$. An augmenting path method for the solution of the polymatroid intersection problem together with complexity bounds is developed by Schönsleben [17] ($R = \mathbb{Z}$).

Recently, Frank [this volume, pp. 97–120] developed an appealing primal–dual method for solving the apparently more general combinatorial optimization problem introduced in [2]. For linear objective functions ($R = \mathbb{Z}$) that method is quasipolynomial, in general, and polynomial, if $0 \leqslant x \leqslant 1$.

On the other side, network flows with values in more general totally ordered algebraic structures are discussed by Hamacher [9]. A general account of optimization in algebraic structures is provided in [18]. In Section 2 we show. that polymatroid intersections with capacities, independent flows and polymatroidal flows are equivalent combinatorial structures. The mutual reductions are of linear complexity, Since Fujishige [7] admitted group-valued vectors we can derive group-valued counterparts of Edmonds (integral) polymatroid intersection theorem. In Section 3 we develop a general strategy for the solution of (1.1) which is based on the assumption of the validity of a local optimality criterium in certain cones. We discuss necessary and sufficient conditions. In part, similar conditions are discussed in [4]. For example, linear functions and certain quotient functions satisfy these conditions. We apply the method to maximum matchings and independent flows. In particular, we develop a negative circuit method for the determination of maximum independent flows with minimum objective function value.

2. Polymatroids and flows

Polymatroids were introduced by Edmonds [1]. The following definition generalizes this concept in a certain algebraic sense and is drawn from [7].

Let E be a nonempty, finite set and let $(R, +, \leqslant)$ be a totally ordered, commutative group with neutral element 0. For example, R may be the additive group of real numbers, rational numbers or integers endowed with the usual order relation. Further examples can be found in [17]. The set of all nonnegative elements in the group is denoted by R_+, i.e., $R_+ := \{a \mid 0 \leqslant a\}$. Let $a \in R$ and $m \in \mathbb{N}$. Then we denote $a + a + \cdots + a$ (m times) by $m \cdot a$ or ma. For $x \in R^E$ we define

$$x(A) := \sum_{e \in A} x(e)$$

for all $A \subseteq E$; in particular $x(\emptyset) := 0$.

A function $\rho : 2^E \to R_+$ is called *β-function* if it satisfies

$$\rho(\emptyset) = 0\,, \tag{2.1}$$

$$A \subseteq B \;\Rightarrow\; \rho(A) \subseteq \rho(B) \quad \text{(Isotonicity)}\,, \tag{2.2}$$

$$\rho(A \cup B) + \rho(A \cap B) \leqslant \rho(A) + \rho(B) \quad \text{(Submodularity)} \tag{2.3}$$

for all $A, B \subseteq E$. If ρ is a β-function, then $\mathbb{P} = \mathbb{P}(E, \rho)$ is called a *polymatroid* with ground set E and *ground set rank function* ρ. A vector $x \in R_+^E$ is called *independent* (with respect to $\mathbb{P}$) if

$$x(A) \leqslant \rho(A) \tag{2.4}$$

for all $A \subseteq E$. The set of all independent vectors is denoted by P. In particular, if R is the additive group of real numbers, then P is a polytope in $\mathbb{R}_+^E$. In the general case, P is a subset of R_+^E which is the intersection of a finite number of 'half-spaces' of R^E (cf. (2.4)). Such algebraic approaches are proposed for many combinatorial optimization problems; a general discussion can be found in [18].

For two polymatroids $\mathbb{P}_1(E, \rho_1)$ and $\mathbb{P}_2(E, \rho_2)$, a vector $x \in P_1 \cap P_2$ is called a (polymatroid) *intersection*. We admit additional upper bounds on intersections. Let $c \in R_{++}^E$ ($R_{++} := R_+ \backslash \{0\}$). Then x is called *feasible* if $x \leqslant c$. The set of all feasible intersections x with $x(E) = \alpha$ for fixed $\alpha \in R_+$ is denoted by $P(\alpha)$, i.e.,

$$P(\alpha) := \{x \in R_+^E \mid x \in P_1 \cap P_2,\; x \leqslant c,\; x(E) = \alpha\}\,. \tag{2.5}$$

The cardinality $|E|$ of E is called the *size of* $P(\alpha)$.

It is well known that polymatroids generalize matroids (cf. [1]). A common generalization of matroids and flows is developed in a series of papers by Iri and Tomizawa [12], Fujishige [5, 6] and Iri [11]. The general algebraic case is considered by Fujishige [7]. He introduces independent flows in the following way.

Let $G = G(V, A, V_1, V_2)$ be a finite digraph with vertex set V, arc set A and $V_1, V_2 \subseteq V$. Further, capacities $c \in R_{++}^A$ and two polymatroids $\mathbb{P}_i(V_i, \rho_i)$, $i = 1, 2$, are given. The set of all arcs into (out of) a subset $V' \subseteq V$ is denoted by $\omega^-(V')$ ($\omega^+(V')$). $x \in R_+^A$ is called a *flow* in G if the *conservation law*

$$x(\omega^-(v)) = x(\omega^+(v)) \tag{2.6}$$

holds for all $v \in V \backslash (V_1 \cup V_2)$. Then we define $s \in R^{V_1}$ and $t \in R^{V_2}$ by

$$\begin{aligned} s(u) &:= x(\omega^+(u)) - x(\omega^-(u))\,, \\ t(w) &:= x(\omega^-(w)) - x(\omega^+(w))\,, \end{aligned} \tag{2.7}$$

for all $u \in V_1$, $w \in V_2$. A flow x is called *independent* (with respect to $\mathbb{P}_1$ and $\mathbb{P}_2$) if $s \in P_1$ and $t \in P_2$. If x is an independent flow, then $s \geq 0$ and $t \geq 0$. Therefore V_1 (V_2) is called the *set of the sources* (*sinks*) of G. A flow x is called *feasible* if $x \leq c$.

Due to (2.6) a flow x satisfies $s(V_1) = t(V_2)$. We call $|x| := s(V_1)$ the *flow value of* x. The set of all feasible independent flows of flow value α for fixed $\alpha \in R_+$ is denoted by $G(\alpha)$, i.e.,

$$G(\alpha) = \{x \in R_+^A \mid x \text{ feasible, independent flow; } |x| = \alpha\}. \tag{2.8}$$

The pair $(|V|, |A|)$ is called the *size of* $G(\alpha)$.

The following reduction shows that independent flows generalize polymatroid intersections.

Let $E' = \{e' \mid e \in E\}$ denote a copy of the ground set E and let $A := \{(e, e') \mid e \in E\}$ (cf. Fig. 1).

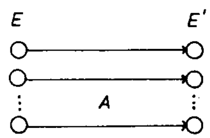


Fig. 1. $P(\alpha) \rightsquigarrow G(\alpha)$.

We define capacities $\tilde{c} \in R_{++}^A$ by $\tilde{c}(e, e') := c(e)$ for all $e \in E$. The polymatroid $\mathbb{P}_2$ is redefined as a polymatroid $\tilde{\mathbb{P}}_2$ on E' with ground set rank function $\tilde{\rho}_2$:

$$\tilde{\rho}_2(U') = \rho_2(\{e \mid e' \in U'\})$$

for all $U' \subseteq E'$. Then a feasible independent flow $\tilde{x}$ in $G(E \cup E', A, E, E')$ with capacities $\tilde{c}$ and polymatroids $\mathbb{P}_1$, $\tilde{\mathbb{P}}_2$ uniquely corresponds to a feasible polymatroid intersection $x \in P_1 \cap P_2$. Here, $\tilde{x}(e, e') = x(e)$ for all $e \in E$ and $|\tilde{x}| = x(E)$. We remark that the described reduction is linear in the size $|E|$ of $P(\alpha)$, since the size of the generated independent flow problem is $(2|E|, |E|)$.

Another common generalization of matroids and flows is considered by Martel [16] and by Lawler [14]. They introduce polymatroidal flows in the following way. Let $N = N(V, A)$ be a digraph with vertex set V, containing a source s and a sink t, and with arc set A. We assume that $\omega^-(s) = \omega^+(t) = \emptyset$. Then, $x \in R_+^A$ is called a *flow* if (2.6) is satisfied for all $v \in V \setminus \{s, t\}$. Therefore, a flow x satisfies $x(\omega^+(s)) = x(\omega^-(t)) =: |x|$. That value $|x|$ is called the *flow value of* x.

Now, for all vertices $j \in V$, two polymatroids $\mathbb{P}_{1j}(\omega^+(j), \alpha_j)$ and $\mathbb{P}_{2j}(\omega^-(j), \beta_j)$ are given. A flow x is called *polymatroidal* if the restrictions of x to $\omega^+(j)$ and

$\omega^-(j)$ are independent in the polymatroids $\mathbb{P}_{1j}$ and $\mathbb{P}_{2j}$ for all $j \in V$, i.e., if

$$x|_{\omega^+(j)} \in P_{1j}, \qquad x|_{\omega^-(j)} \in P_{2j},$$

for all $j \in V$. In addition to the concept in [14, 16] we admit capacities $c \in R^A_{++}$. Then, a flow x is called *feasible* if $x \leqslant c$. The set of all feasible polymatroidal flows of value α ($\alpha \in R_+$) is denoted by $N(\alpha)$, i.e.,

$$N(\alpha) := \{x \in R^A_+ \mid x \text{ feasible polymatroidal flow}, |x| = \alpha\}. \tag{2.9}$$

We call $(|V|, |A|)$ the *size of* $N(\alpha)$.

We show that independent flows may be reduced to polymatroidal flows. We adjoin a new source s, a new sink t and the arc sets $\{s\} \times V_1$, $V_2 \times \{t\}$ to the underlying digraph G (cf. Fig. 2). In this way we get the digraph $N = N(V \cup \{s, t\}, A^*)$ with $A^* = A \cup (\{s\} \times V_1) \cup (V_2 \times \{t\})$. All new arcs have capacity α (w.l.o.g. $\alpha > 0$). We define two polymatroids $\mathbb{P}_{1s}(\omega^+(s), \alpha_s)$ and $\mathbb{P}_{2t}(\omega^-(t), \beta_t)$ by

$$\alpha_s(X) := \rho_1(\{u \mid (s, u) \in X\}),$$

$$\beta_t(Y) := \rho_2(\{w \mid (w, t) \in Y\})$$

for all $X \subseteq \omega^+(s)$ and for all $Y \subseteq \omega^-(t)$. Further polymatroids are not necessary. For sake of completeness, let $D \in R_+$ be an upper bound on the value of the total amount of flow entering or leaving a vertex in G, for example, $D = \Sigma_{a \in A}\, c(a)$. Then all polymatroids are defined with ground set rank function

$$\rho(X) := \begin{cases} 0 & \text{if } X = \emptyset \\ D & \text{otherwise}, \end{cases} \tag{2.10}$$

for all subsets X of the considered ground set.

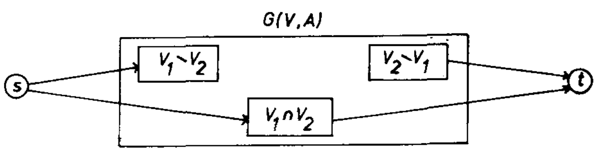


Fig. 2. $G(\alpha) \rightsquigarrow N(\alpha)$.

Now, a feasible independent flow $x \in G(\alpha)$ induces a feasible polymatroidal flow $\bar{x} \in N(\alpha)$, where

$$\bar{x}(s, u) := s(u) = x(\omega^+(u)) - x(\omega^-(u)),$$

$$\bar{x}(w, t) := t(w) = x(\omega^-(w)) - x(\omega^+(w))$$

for all $u \in V_1$, $w \in V_2$. Clearly, the restriction x of a feasible polymatroidal flow $\tilde{x} \in N(\alpha)$ to A is a feasible independent flow $x \in G(\alpha)$. We remark that this reduction is linear in the size of $G(\alpha)$; in fact, the size of the generated problem $N(\alpha)$ is $(|V|+2, |A|+|V_1|+|V_2|)$.

The above reductions are rather straightforward; a more surprising result is contained in the following theorem.

Theorem 2.1. *Let $\alpha \in R_+$. Then*
(1) $P(\alpha)$ can be reduced to some $G(\alpha)$,
(2) $G(\alpha)$ can be reduced to some $N(\alpha)$,
(3) $N(\alpha)$ can be reduced to some $P(\alpha + H)$
where $H \in R_+$ is sufficiently large. In fact, $H \leqslant (n-2) \cdot D$ where n is the number of vertices in N and D is the sum of all capacities on N.

Proof. We have already described the reductions of $P(\alpha)$ to $G(\alpha)$ and of $G(\alpha)$ to $N(\alpha)$. Therefore it suffices to describe a reduction of $N(\alpha)$ to some $P(\alpha + H)$ which is linear in the size of $N(\alpha)$.

Let $D := \Sigma_{a \in A}\, c(a)$ and let

$$d_j := \min(\alpha_j(\omega^+(j)), \beta_j(\omega^-(j)), D) \tag{2.11}$$

for $j \in V_0 := V \setminus \{s, t\}$. Let $H := \Sigma_j\, d_j$.

In the first step we reduce $N(\alpha)$ to a bipartite transportation problem with additional polymatroidal constraints. This step is inspired by a well-known transformation of the classical transshipment problem to the transportation problem (cf. [13]). We split every vertex $j \in V_0$ into two vertices u_j and w_j. The new bipartite digraph $T = (U, W, A')$ has vertex sets $U = \{u_j \mid j \in V_0 \cup \{s\}\}$ and $W = \{w_j \mid j \in V_0 \cup \{t\}\}$. The new arc set $A' = A_1 \cup A_2$ consists of $A_1 := \{(u_j, w_j)_0 \mid j \in V_0\}$, and of all arcs of the form

$$(u_i, w_j)_k, \quad k = 1, 2, \ldots, k_{ij}, \tag{2.12}$$

if the original digraph N contained the arcs

$$(i, j)_k, \quad k = 1, 2, \ldots, k_{ij},$$

for $i \in V_0 \cup \{s\}$, $j \in V_0 \cup \{t\}$ (cf. Fig. 3).

We assign capacity D to all arcs in A_1. The capacity of an arc $(u_i, w_j)_k \in A_2$ is the same as the capacity of its corresponding arc $(i, j)_k \in A$. The polymatroids $\mathbb{P}_{1j}$ $(j \neq t)$ and $\mathbb{P}_{2j}$ $(j \neq s)$ for $j \in V$ lead to the following corresponding

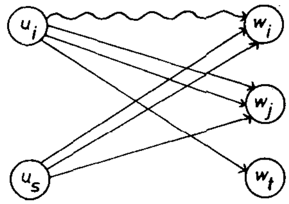


Fig. 3. $N(\alpha) \rightsquigarrow T$. Wavy arcs ($\in A_1$) are not present in the original digraph N.

polymatroids. For $X \subseteq A'$, let $\bar{X} \subseteq A$ denote the set of all arcs in A corresponding to $X \backslash A_1$, i.e.,

$$\bar{X} := \{(i, j)_k \in A \mid (u_i, w_j)_k \in X, k > 0\}.$$

Let $j \in V_0$ and $X \subseteq \omega^+(u_j)$. We define

$$\alpha'_{u_j}(X) := \begin{cases} \min(d_j, \alpha_j(\bar{X})) & \text{if } (u_j, w_j)_0 \notin X, \\ d_j & \text{otherwise}. \end{cases} \tag{2.13}$$

Then $\mathbb{P}_{1u_j} = \mathbb{P}_{1u_j}(\omega^+(u_j), \alpha'_{u_j})$ is a polymatroid. For $X \subseteq \omega^+(u_s)$, we define

$$\alpha'_{u_s}(X) := \min(\alpha, \alpha_s(\bar{X})).$$

Then $\mathbb{P}_{1u_s} = \mathbb{P}_{1u_s}(\omega^+(u_s), \alpha'_{u_s})$ is a polymatroid.

Similarly, let $w_j \in W$ and $Y \subseteq \omega^-(w_j)$. Then

$$\beta'_{w_j}(Y) := \begin{cases} \min(d_j, \beta_j(\bar{Y})) & \text{if } (u_j, w_j)_0 \notin Y, \\ d_j & \text{otherwise}, \end{cases} \tag{2.14}$$

defines a polymatroid $\mathbb{P}_{2w_j} = \mathbb{P}_{2w_j}(\omega^-(w_j), \beta'_{w_j})$. For $Y \subseteq \omega^-(w_t)$ we define

$$\beta'_{w_t}(Y) := \min(\alpha, \beta_t(\bar{Y})).$$

Then $\mathbb{P}_{2w_t}(\omega^-(w_t), \beta'_{w_t})$ is a polymatroid.

A flow $x \in N(\alpha)$ uniquely corresponds to a transportation flow x' of maximum flow value $x'(A') = \alpha + H$ in T where the vertices in U and V act as multiple sources and sinks. For given $x \in N(\alpha)$ we define x' by

$$x'((u_i, w_j)_k) := \begin{cases} x((i, j)_k) & \text{if } k > 0, \\ d_i - x(\omega^+(i)) & \text{otherwise} \end{cases}$$

for all arcs in A'. Due to the conservation law

$$x(\omega^+(i)) = x(\omega^-(i))$$

for all $i \in V_0$. Therefore, x' satisfies all constraints and has maximum flow value $x'(A') = \alpha + H$. Vice versa, let x' be a transportation flow of value $\alpha + H$. Due to the polymatroidal constraints we find

$$x'(\omega^+(u_i)) = x'(\omega^-(w_i)) = d_i,$$

$$x'(\omega^+(u_s)) = x'(\omega^-(w_t)) = \alpha$$

for all $i \in V_0$. Therefore,

$$x((i,j)_k) := x'((u_i, w_j)_k)$$

for all arcs in A defines a feasible, polymatroidal flow of value α in N.

In the second step of our reduction we observe that a transportation flow $x' \in R_+^{A'}$ can be interpreted as an intersection of two polymatroids defined on A' as their common ground set, and vice versa. Let $X \subseteq X'$. We define

$$\rho_1(X) := \sum_{i \in V_0} \alpha'_{u_i}(X \cap \omega^+(u_i)) + \alpha'_{u_s}(X \cap \omega^+(u_s)), \tag{2.15}$$

$$\rho_2(X) := \sum_{i \in V_0} \beta'_{w_i}(X \cap \omega^-(w_i)) + \beta'_{w_t}(X \cap \omega^-(w_t)). \tag{2.16}$$

Then $\mathbb{P}_k(A', \rho_k)$, $k = 1, 2$, are polymatroids. Clearly, x' is a feasible, polymatroidal transportation flow of maximum flow value $x'(A') = \alpha + H$ iff x' is a feasible intersection of $\mathbb{P}_1$ and $\mathbb{P}_2$ of maximum component sum $\alpha + H$.

Therefore, a complete reduction of $N(\alpha)$ to $P(\alpha + H)$ is found which is linear in the size of $N(\alpha)$. In fact, the size of the generated problem is $|A'| = |A| + |V_0|$. □

Theorem 2.1 shows that polymatroid intersections, independent flows and polymatroidal flows define equivalent combinatorial structures which can be reduced to each other by the described transformations. $P(\alpha)$ or the corresponding bipartite version of $G(\alpha)$ seem to be the simpler representations of that structure. Nevertheless, it might be interesting to see how well-known results are reflected by the different choices of the representation.

Theorem 2.2 (Fujishige [7]). *Let F denote the set of all feasible, independent flows in $G = G(V, A, V_1, V_2)$. Then*

(1) *there exists $\tilde{x} \in F$ such that*

$$|\tilde{x}| = \sup\{|x| \mid x \in F\},$$

(2) $|\tilde{x}| = \min\{\rho_1(V_1 \backslash U) + c(A \cap (U \times (V \backslash U))) + \rho_2(V_2 \cap U) \mid U \subseteq V\}$.

We remark that, in the particular case $R = \mathbb{Z}$, the existence of an integer valued feasible, independent flow is claimed implicitly. Let $S, T \subseteq V$ with $S \cup T = V$, $S \cap T = \emptyset$. Then (S, T) is called a *partition of V*.

An equivalent formulation of Theorem 2.2 is the following theorem.

Theorem 2.3. *Let $\alpha \in R_+$. Then $G(\alpha) \neq \emptyset$ if and only if*

$$\rho_1(V_1 \cap T) + c(A \cap (S \times T)) + \rho_2(V_2 \cap S) \geqslant \alpha$$

for all partitions (S, T) of V.

The equivalent theorem in terms of $P(\alpha)$ can be derived from inspection of Fig. 1 which shows the digraph $G(\alpha)$ corresponding to a given $P(\alpha)$.

Theorem 2.4. *Let $\alpha \in R_+$. Then $P(\alpha) \neq \emptyset$ if and only if*

$$\rho_1(X) + c(E \backslash (X \cup Y)) + \rho_2(Y) \geqslant \alpha$$

for all $X, Y \subseteq E$ with $X \cap Y = \emptyset$.

Proof. Let $G = G(E \cup E', A, E, E')$ be the digraph in Fig. 1. Then, by Theorem 2.3, $G(\alpha) \neq \emptyset$ if and only if

$$\rho_1(E \cap T) + \tilde{c}(A \cap (S \times T)) + \rho_2(E' \cap S) \geqslant \alpha$$

for all partitions (S, T) of $E \cup E'$. The claimed result follows from Fig. 4. □

Theorem 2.4 implies polymatroid intersection theorems from Edmonds [1], Giles [8] and others. We emphasize that the case $R = \mathbb{Z}$ describes the case of integral polymatroids.

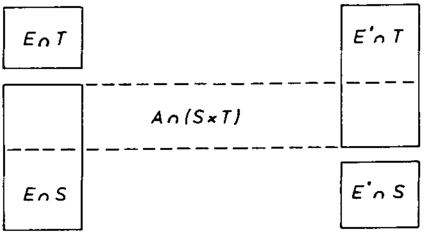


Fig. 4. Theorem 2.3 ⇒ Theorem 2.4.

Analyzing the reduction in Theorem 2.1 we find the following equivalent theorem in terms of $N(\alpha)$.

We call a partition (S, T) of V a *cut of* $N = N(V, A)$ if $s \in S$ and $t \in T$.

Theorem 2.5. *Let $\alpha \in R_+$. Then $N(\alpha) \neq \emptyset$ if and only if*

$$\sum_{v\in S} \alpha_v(U \cap \omega^+(v)) + \sum_{v \in T} \beta_v(L \cap \omega^-(v)) + c[(A \cap S \times T)\backslash(U \cup L)] \geqslant \alpha$$

for all cuts (S, T) and for all disjoint $U, L \subseteq A \cap S \times T$.

Proof. In the final step of Theorem 2.1 we obtain two polymatroids $\mathbb{P}_k(A', \rho_k)$, $k = 1, 2$. We identify A_1 with V_0 and A_2 with A ($A' = A_1 \cup A_2$) and we denote $V_0 \cup A$ by E. In particular, we get polymatroids $\mathbb{P}_1(E, \rho_1)$, $\mathbb{P}_2(E, \rho_2)$ and capacities $c^* \in R^E_{++}$. Theorem 2.4 shows that $N(\alpha) \neq \emptyset$ if and only if

$$\rho_1(X) + \rho_2(Y) + c^*(E\backslash(X \cup Y)) \geqslant \alpha + \sum_{j \in V_0} \alpha_j$$

for all $X, Y \subseteq E$ with $X \cap Y = \emptyset$.

For $x \subseteq E$ let $X_0 := X \cap V_0$ and $X^\pm_v := X \cap \omega^\pm(v)$ for all $v \in V$. Replacing d_j, $j \in V_0$, by a sufficiently large constant d, the definition of ρ_1 (cf. (2.13) and (2.15)) leads to

$$\rho_1(X) = |X_0| \cdot d + \sum_{v \notin X_0} \alpha_v(X^+_v) + \min(\alpha_s(X^+_s), \alpha)\,.$$

The similar expression for ρ_2 (cf. (2.14) and (2.16)) is

$$\rho_2(Y) = |Y_0| \cdot d + \sum_{v \notin Y_0} \beta_v(Y^-_v) + \min(\beta_t(Y^-_t), \alpha)\,.$$

for $Y \subseteq E$. The capacities are

$$c^*(E\backslash(X \cup Y)) = (|V_0| - |X_0 \cup Y_0|) \cdot d + c(A\backslash(X_A \cup Y_A))\,,$$

with $X_A = X \cap A$, $Y_A = Y \cap A$ and the capacity function c of $N(V, A)$.

It suffices to consider X, Y such that $X_0 \cup Y_0 = V_0$; since $X_0 \cap Y_0 = \emptyset$ we find that (S, T) is a cut for

$$S := (V_0\backslash X_0) \cup \{s\}\,, \qquad T := (V_0\backslash Y_0) \cup \{t\}\,.$$

Since $X_0 \cap Y_0 = \emptyset$, the d-terms cancel on both sides of the inequalities. The validity of the inequality is not changed if we replace the minima $\min(\alpha_s(X_s^+), \alpha)$, $\min(\beta_t(Y_t^-), \alpha)$ by $\alpha_s(X_s^+)$, $\beta_t(X_t^-)$. Therefore $N(\alpha) \neq \emptyset$ if and only if

$$\sum_{v \in S} \alpha_v(X_v^+) + \sum_{v \in T} \beta_v(Y_v^-) + c(A \setminus (X \cup Y)) \geqslant \alpha$$

for all cuts (S, T) and for all $X, Y \subseteq A$. The choice of X and Y can be restricted in the following way. If an arc from $A \setminus S \times T$ does not belong to $X \cup Y$, then it can be added to X or Y without enlarging the sets X_v^+, Y_v^-. Further, if such an arc belongs to some X_v^+ then it may be shifted to Y without enlarging the set Y_v^-, and vice versa. Such changes will not increase the left hand side of the considered inequality. Thus, the arcs of $A \setminus (S \times T)$ can be excluded from the considered inequalities. The relevant part of X, Y is called U, L. $\square$

In particular, Theorem 2.5 shows the existence of an integral feasible polymatroidal flow of maximum value in the case $R = \mathbb{Z}$. The three theorems have the following corollary which describes the case without active capacities.

If (S, T) is a cut in $N = N(V, A)$ and (U, L) is a partition of $A \cap (S \times T)$, then (S, T, U, L) is called an *arc-partitioned cut of* N (cf. [15]).

Corollary 2.6. *Let* $\alpha \in R_+$. *Then*

(1) $P_\alpha := \{x \in R_+^E \mid x$ *intersection in* $P_1 \cap P_2$, $x(E) = \alpha\} \neq \emptyset$ *iff*

$$\rho_1(X) + \rho_2(E \setminus X) \geqslant \alpha$$

for all $X \subseteq E$,

(2) $G_\alpha := \{x \in R_+^A \mid x$ *independent flow in* G, $|x| = \alpha\} \neq \emptyset$ *iff*

$$\rho_1(V_1 \cap T) + \rho_2(V_2 \cap S) \geqslant \alpha$$

for all partitions (S, T) *of* V *with* $A \cap (S \times T) = \emptyset$,

(3) $N_\alpha := \{x \in R_+^A \mid x$ *polymatroidal flow in* N, $|x| = \alpha\} \neq \emptyset$ *iff*

$$\sum_{v \in S} \alpha_v(U \cap \omega^+(v)) + \sum_{v \in T} \beta_v(L \cap \omega^-(v)) \geqslant \alpha$$

for all arc-partitioned cuts (S, T, U, L).

In particular, Corollary 2.6(3) implies the max-flow min-cut theorems in [14] ($R = \mathbb{Z}$, $\mathbb{Q}$) and, recently, in [15] ($R = \mathbb{R}$).

For the determination of an intersection with maximum component sum $\alpha_{\max}$ Schönsleben [17] develops a polynomially bounded augmenting path method ($R = \mathbb{Z}$, without capacities). He provides a complexity bound $O(\min(n^3, \kappa) \cdot Z \cdot n^2(\log \kappa + n))$ where $n = |E|$, $\kappa := \min\{k \in \mathbb{N} \mid P_1 \cap P_2 \subseteq [0, k]^E\}$, and where the complexity of an independence oracle in P_1 and P_2 is Z. Clearly, using the previously described reductions, this method can be transformed to a polynomially bounded augmenting path method for the determination of independent network flows/polymatroidal network flows of maximum flow value $\alpha_{\max}$. If $\alpha_{\max}$ is known, the method can directly be applied to the respectively derived intersection problem ($\alpha = \alpha_{\max}$). Otherwise, we have to take care of the fact that the rank functions of that intersection problem (cf. (2.15) and (2.16)) depend on α. A simple approach consists in a combination of the augmenting path method with binary search for $\alpha_{\max}$.

Recently, Lawler and Martel [15] derived an augmenting path method for the determination of polymatroidal network flows of maximum flow value ($R \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$, without capacities). They show that no more than $O(n^3)$ augmentations are required and discuss detailed complexity bounds with respect to certain oracles used within each augmentation. Clearly, using the respective reductions, that method applies to intersections/independent flows.

Recently, Frank [3] developed an appealing primal–dual method for solving a weighted combinatorial optimization problem introduced in [2]. In particular, that problem generalizes intersections, independent network flows and polymatroidal network flows. Therefore Frank's method applies to the determination of 'maximum' intersections/independent network flows/polymatroidal network flows ($R = \mathbb{Z}$). The method is quasipolynomial, in general, and polynomial, if $0 \leqslant x \leqslant 1$.

Fujishige [7] develops an augmenting path method for the determination of independent flows of minimum weight in totally ordered, commutative rings, which, in particular, is applicable to the determination of independent flows of maximum flow value. Then, the method is valid in the general case of totally ordered, commutative groups $(R, +, \leqslant)$ but finite termination is assured only if $R \in \{\mathbb{Z}, \mathbb{Q}\}$. Fujishige [7] uses a certain auxiliary graph generalizing the border-graph of Krogdahl (cf. [13]) as well as the usual incremental graph in flow theory. That graph is a basic tool for the further development of a primal solution method for the minimization of linear objective functions in the case of totally ordered, commutative rings. We will generalize that method in Section 3 for certain nonlinear objective functions.

Before describing the auxiliary graph we need some preliminaries from polymatroid theory. Let $\mathbb{P}(E, \rho)$ denote a polymatroid. The *saturation* function $\operatorname{sat} : P \to 2^E$ is defined by

$$\mathrm{sat}(x) := \{i \in E \mid x + \varepsilon^i \notin P \; \forall \varepsilon > 0\}$$

for $x \in P$ ($\varepsilon^i \in R_+^E$ is defined by $\varepsilon_i^i := \varepsilon$; $\varepsilon_j^i := 0$ for all $j \neq i$). We remark that $i \in \mathrm{sat}(x)$ iff $x(A) = \rho(A)$ for some $A \subseteq E$ with $i \in A$. Let

$$c^+(x, i) := \max_{\varepsilon \geq 0}\{\varepsilon \mid x + \varepsilon^i \in P\} = \min_{A \subseteq E}\{\rho(A) - x(A) \mid i \in A\}$$

denote the (unsaturated) gap in the ith component of x. Clearly, $c^+(x, i) \geq 0$, and $i \notin \mathrm{sat}(x)$ iff $c^+(x, i) > 0$. The *dependence function* $\mathrm{dep}: P \times E \to 2^E$ is defined by $\mathrm{dep}(x, i) = \emptyset$, if $i \notin \mathrm{sat}(x)$, and by

$$\mathrm{dep}(x, i) := \{j \in \mathrm{sat}(x) \mid \exists \varepsilon > 0 \colon x + \varepsilon^i - \varepsilon^j \in P\},$$

if $i \in \mathrm{sat}(x)$. Let

$$c'(x, i, j) = \max_{\varepsilon \geq 0}\{\varepsilon \mid x + \varepsilon^i - \varepsilon^j \in P\},$$

for $x \in P$ and $j \in \mathrm{dep}(x, i) \backslash \{i\}$. With

$$\gamma := \min_{A \subseteq E}\{\rho(A) - x(A) \mid i \in A, j \notin A\}$$

we find

$$0 < c'(x, i, j) = \min\{\gamma, x(j)\}.$$

In view of Theorem 2.1 it suffices to consider the auxiliary graph for the bipartite digraph $G(E \cup E', A, E, E')$ corresponding to polymatroid intersections (cf. Fig. 1). We remind that a capacity function $c \in R_{++}^A$ and two polymatroids $\mathbb{P}_1(E, \rho_1)$ and $\mathbb{P}_2(E', \rho_2)$ are given. Let x be a feasible independent flow in G. s, t are defined by $s(i) = x(i, i') = t(i')$ for all $e \in E$. Then the auxiliary digraph $G_x(V \cup \{\sigma, \tau\}, \tilde{A})$ contains six disjoint arcs sets (cf. Fig. 5).

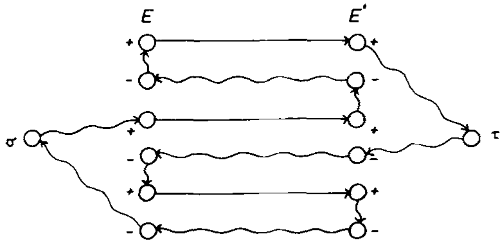


Fig. 5. Circuit in auxiliary graph G_x. Wavy arcs are not from A. Signs indicate increase (+) and decrease (−) in (x, s, t) due to a flow on the circuit.

The first two sets are drawn from the incremental graph in flow theory:

$$F := \{(i, i') \mid x(i, i') < c(i, i')\} \quad \text{'forward arcs'},$$
$$B := \{(i', i) \mid x(i, i') > 0\} \quad \text{'backward arcs'}.$$

Secondly, two sets generalize the dependence graph of matroids for the polymatroid $\mathbb{P}_1(E, \rho_1)$:

$$D_1 := \{(j, i) \mid i \in \mathrm{sat}_1(s), j \in \mathrm{dep}_1(s, i)\backslash\{i\}\},$$
$$S_1 := \{(\sigma, i) \mid i \in E\backslash\mathrm{sat}_1(s)\} \cup \{(i, \sigma) \mid s(i) > 0\}.$$

Thirdly, a similar dependence graph is defined for $\mathbb{P}_2(E', \rho_2)$ with reversed arc directions:

$$D_2 := \{(i', j') \mid i' \in \mathrm{sat}_2(t), j' \in \mathrm{dep}_2(s, i')\backslash\{i'\}\},$$
$$S_2 := \{(i', \tau) \mid i' \in E'\backslash\mathrm{sat}_2(t)\} \cup \{(\tau, i') \mid t(i') > 0\}.$$

We assign a positive capacity $\tilde{c}$ to all arcs in $\tilde{A}$:

$$\begin{aligned}
\tilde{c}(i, i') &:= c(i, i') - x(i, i'), && (i, i') \in F,\\
\tilde{c}(i', i) &:= x(i, i'), && (i', i) \in B,\\
\tilde{c}(j, i) &:= c_1'(s, i, j), && (j, i) \in D_1,\\
\tilde{c}(\sigma, i) &:= c_1^{\dagger}(s, i), && (\sigma, i) \in S_1,\\
\tilde{c}(i, \sigma) &:= s(i), && (i, \sigma) \in S_1,\\
\tilde{c}(i', j') &:= c_2'(t, i', j'), && (i', j') \in D_2,\\
\tilde{c}(i', \tau) &:= c_2^{\dagger}(t, i'), && (i', \tau) \in S_2,\\
\tilde{c}(\tau, i') &:= t(i'), && (\tau, i') \in S_2.
\end{aligned}$$

A flow Δx in the network G_x with source σ and sink τ is called *F-B-feasible* if $\Delta x(i, i') \leqslant \tilde{c}(i, i')$ for all $(i, i') \in F$ and $\Delta x(i', i) \leqslant \tilde{c}(i', i)$ for all $(i', i) \in B$. An F-B-feasible flow Δx defines a feasible flow $\bar{x}$ in G by

$$\bar{x}(i, i') := (x \oplus \Delta x)(i, i') := x(i, i') + \Delta x(i, i') - \Delta x(i', i)$$

for all $i \in E$ (we interpret Δx by 0 if an arc does not occur in G_x). The following theorem is implicitly proved in [7].

Theorem 2.7. *Let $\bar{x} \in G(\beta)$, $x \in G(\alpha)$ with $\beta \geqslant \alpha$. Then there exists an F-B-feasible flow Δx in G_x of flow value $\beta - \alpha$ such that $\bar{x} = x \oplus \Delta x$.*

From usual flow theory we know that a flow can be decomposed in *path flows* and *circuit flows*, i.e., in flows which have a positive constant value on the arcs of a fixed direct path from source to sink (or of a fixed directed circuit), and which have value 0 on all other arcs in the network. Applied to Δx in Theorem 2.7 this shows $\Delta x = \Delta x_p + \Delta x_c$ where Δx_p is the sum of certain path flows and Δx_c is the sum of certain circuit flows. In particular, Δx_p has flow value $\beta - \alpha$ and, if $\beta = \alpha$, then $\Delta x = \Delta x_c$.

On the other hand, different from usual flow theory it may happen that $x \oplus \Delta x$ is not an independent flow in G even if Δx is a feasible flow in G_x. For feasible path (circuit) flows Δx Fujishige [7] describes a sufficient criterion for the independence of $x \oplus \Delta x$.

It suffices to consider path (circuit) flows Δx with the following property (cf. Fig. 5):

Any vertex ($\neq \sigma, \tau$) of the path (circuit) is incident to an arc of the path which belongs to F and B. (2.17)

If Δx does not satisfy (2.17), then it is easy to construct a path (circuit) flow $\Delta x'$ satisfying (2.17) such that $x \oplus \Delta x = x \oplus \Delta x'$. This construction consists in the introduction of some shortcutting arcs from D_i, S_i ($i = 1, 2$) which circumvent the unnecessary vertices.

Thus, the arcs of the path (circuit) which belong to D_1 or D_2 have pairwise disjoint end-vertices. These arcsets are denoted by Δ_1 and Δ_2. We define digraphs $G_\mu(\Delta_\mu, A_\mu)$ where the arcset A_μ consists in all arcs $[(i, j), (k, r)]$ such that $(i, r) \in D_\mu$ ($\mu = 1, 2$). If Δx satisfies (2.17) and neither G_1 nor G_2 does contain a directed circuit, then Δx is called *admissible*.

The following result is drawn from Fujishige [7].

Theorem 2.8. *Let $x \in G(\alpha)$ and let Δx be an admissible, feasible path (circuit) flow in G_x with flow value δ (clearly, $\delta = 0$ in case of a circuit). Then $x \oplus \Delta x \in G(\alpha + \delta)$.*

Theorems 2.7 and 2.8 directly lead to an augmenting path method for the construction of an independent flow of maximum flow value. If $x \in G(\alpha)$ is known but α is not maximal, then there exists a directed path from σ to τ in G_x. We choose a path of shortest length and push as much flow through this path as possible without violation of feasibility; the resulting flow Δx is admissible (no shortcuts possible). The procedure is finite, at least if $R = \mathbb{Z}$ or $R = \mathbb{Q}$.

3. Minimization of objective functions

In this section $(R, +, \cdot, \leq)$ is a totally ordered, commutative ring with zero 0 and unity 1. Thus, in particular, $(R, +, \leq)$ is a totally ordered, commutative group as in Section 2. We discuss the minimization of an objective function $f: D \to T$ where $D \subseteq R^n$ and where $(T, \leq)$ is a totally ordered set. We use the same notation '$\leq$' for the order relations in R and T but from the context it will always be clear which is meant specifically. A typical example is a linear function $f(x) = c^T x$ with $c \in R^n$ and $R = T \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$. The set of feasible solutions is denoted by P ($\subseteq D$). Thus, the minimization problem reads

$$\min\{f(x) \mid x \in P\}. \tag{3.1}$$

The strategy developed for solving (3.1) may be modified for maximization problems in a straightforward way.

Let $x, y \in R^n$. Then

$$[x, y] := \{x + \lambda(y - x) \mid 0 \leq \lambda \leq 1, \lambda \in R\}.$$

Let $S \subseteq R^n$. Then S is called *convex* with respect to R, if $[x, y] \subseteq S$ for all $x, y \in S$. For $x \in R^n$ we define $S - x := \{y - x \mid y \in S\}$. The *convex cone generated by S* with respect to R consists in all nonnegative, finite linear combinations of the elements in S, i.e.,

$$\operatorname{cone}_R(S) := \left\{\sum_{i=1}^{m} \lambda_i x^i \mid x^i \in S, \lambda_i \in R_+; m \in N\right\}.$$

We will only consider functions satisfying the following property.

Property 3.1. *Let $x \in D$ and $S \subseteq D$. If $f(x) \leq f(y)$ for all $y \in S$, then $f(x) \leq f(y)$ for all $y \in [x + \operatorname{cone}_R(S - x)] \cap D$.*

Property 3.1 enables the use of local optimality criteria in order to prove global optimality in certain cones. Therefore the following procedure is proposed for solving (3.1):

General strategy for solving (3.1) ($P' \subseteq P$)
Step 1. Find $x \in P'$.
Step 2. If x is locally optimal (in $Q = Q(x)$) stop.
Step 3. Otherwise find $y \in Q(x)$ with $f(y) < f(x)$; $x := y$ and go to Step 2.

Clearly, the strategy is valid if we can find sets $Q(x)$ (for $x \in P'$) such that

$$Q(x) \subseteq P' \quad \text{and} \quad P \subseteq x + \text{cone}_R(Q(x) - x) \tag{3.2}$$

for all $x \in P'$. Finiteness of the strategy will follow from trivial arguments for the considered problems (e.g. P' finite).

For example, let $R = \mathbb{R}$ and let P be a polytope. Then, linear functions satisfy Property 3.1 ($D = \mathbb{R}^n$). For a vertex $x \in P$, the set $Q(x)$ of all vertices of P adjacent to x satisfies (3.2). The general strategy is a rough description of the primal simplex method. We observe that an arbitrary function satisfying Property 3.1 will attain its minimum over a polytope at a vertex of the polytope ($P' \triangleq$ set of all vertices).

A necessary condition for Property 3.1 is

$$f(x) \leq f(y), f(x) \leq f(z) \Rightarrow f(x) \leq f(x + (y - x) + (z - x)) \tag{3.3}$$

for all x, y, $z \in D$ with $y + z - x \in D$. Let $R = \mathbb{Z}$. Then, if $D = \mathbb{Z}^n$, (3.3) and Property 3.1 are equivalent. In the case $D \subsetneq \mathbb{Z}^n$, (3.3) does not necessarily imply Property 3.1. Another necessary condition for Property 3.1 is

$$f(x) \leq f(y) \Rightarrow f(x) \leq f(x + \lambda(y - x)) \tag{3.4}$$

for all x, $y \in D$, $\lambda \in R_+$ with $x + \lambda(y - x) \in D$. For fields R (3.3) and (3.4) are often sufficient.

Theorem 3.2. *Let R be a field and let $D \subseteq R^n$ be a convex set. If $f : D \to T$ satisfies* (3.3) *and* (3.4), *then f has Property* 3.1.

Proof. Let $y^1, y^2, \ldots, y^m \in S \subseteq D$. Due to convexity

$$y = y(\lambda) = x + \sum \lambda_i(y^i - x) \in D$$

for all $\lambda \in R_+^m$, $\Sigma \lambda_i \leq 1$. Therefore, by induction, (3.3) and (3.4) lead to $f(x) \leq f(y)$. Let $y(\mu) \in D$ for $\mu \in R_+^m$ and let $\gamma := \Sigma \mu_i > 1$. Then $f(y((1/\gamma) \cdot \mu)) \geq f(x)$. Now $x + \gamma(y((1/\gamma) \cdot \mu) - x) = y(\mu)$. Thus, (3.4) implies $f(x) \leq f(y(\mu))$. □

We remark that in Archimedean fields, i.e., if R is a subfield of the real numbers, (3.4) in Theorem 3.2 may be weakened to *quasiconcavity*, i.e.,

$$f(x) \leq f(y) \Rightarrow f(x) \leq f(z) \quad \forall z \in [x, y] \tag{3.4$'$}$$

for all x, $y \in D$.

In the case $R = \mathbb{Z}$, a result similar to Theorem 3.2 is not known. Even if $D = D' \cap \mathbb{Z}^n$ where D' is a set convex with respect to $\mathbb{R}$, counterexamples show that (3.3) does not imply Property 3.1, in general (only for $|S| = 2$ the implication holds).

The following functions satisfy (3.3) and (3.4).

Linear functions: $f : \mathbb{R}^n \to \mathbb{R}$, defined by $f(x) = c^T x + \alpha$ with $c \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$.

Linear quotient functions: $f : D \to \mathbb{R}$ defined by

$$f(x) = \frac{c^T x + \alpha}{d^T x + \beta},$$

with c, $d \in \mathbb{R}^n$, α, $\beta \in \mathbb{R}$ and with $D := \{x \in \mathbb{R}^n \mid d^T x + \beta > 0\}$.

Time-cost objective functions: $f : \mathbb{R}^n_+ \to \mathbb{R}^2$ with $f(x) = (d(x), c(x))$ defined by

$$d(x) = \max\{d_j \mid x_j > 0\}, \qquad c(x) = \sum_{d_j = d(x)} x_j$$

for $d \in \mathbb{R}^n$. Here, $\mathbb{R}^2$ is totally ordered with respect to the usual lexicographic order relation. We remark that $d : \mathbb{R}^n_+ \to \mathbb{R}$ does not satisfy (3.3), in general.

Multicriteria functions: $f : \mathbb{R}^n \to \mathbb{R}^m$, *if* $f_i : \mathbb{R}^n \to \mathbb{R}$ $(1 \leq i \leq m)$ satisfies (3.3) and (3.4) in a stronger version, i.e., when one strict inequality in the assumption implies a strict inequality, respectively. Again, $\mathbb{R}^m$ is lexicographically ordered.

We observe that the above described objective functions attain their minimum over a given polytope P at a vertex of P. An optimum solution can be obtained using the proposed general strategy.

If the set P of all feasible solutions is of combinatorial structure then this structure might be helpful in the development of a procedure from the proposed general strategy. In the following we discuss matching problems and independent network flows. We will always assume that f has Property 3.1.

At first, let P denote the set of all 0–1 incidence vectors of the maximum matchings in an undirected graph $G = G(V, E)$ ($M \subseteq E$ is a *matching* if all its edges have mutually different end vertices). Let $M \subseteq E$ be a matching of maximum cardinality in G ($|M| = m$). A cycle C in G whose arcs alternately belong to and not to M is called an *alternating cycle*. We denote the set of all alternating cycles by $\tilde{C}$. If $C \in \tilde{C}$, then the symmetric difference $M \oplus C$ is again a maximum matching in G. It is well-known that any maximum matching can be obtained in the form

$$M \oplus (C_1 \cup C_2 \cup \cdots \cup C_k) \tag{3.5}$$

where $C_1, C_2, \ldots, C_k$ are mutually disjoint alternating cycles ($k \in \mathbb{N}$). Let $x(M)$ denote the incidence vector of a maximum matching M. We define

$$Q(x(M)) := \{x(M \oplus C) \mid C \in \tilde{C}\}.$$

We observe $Q(x(M)) \subseteq P$ and, due to (3.5), we get $P \subseteq x + \text{cone}_Z(Q(x) - x)$ for all $x \in P$. Therefore, we may apply the proposed general strategy. We call $C \in \tilde{C}$ *negative* if

$$f(x(M \oplus C)) < f(x(M)).$$

This notation is motivated by the case of linear functions where $f(x(M \oplus C)) - f(x(M))$ is the usual weight (negative !) of the considered cycle. The proposed general strategy mainly consists in the detection of negative alternating cycles. It is valid and finite. Since $P \subseteq \mathbb{Z}^n$ we are only interested in the restriction of f to $\mathbb{Z}^n$. It suffices that $f|_{D \cap \mathbb{Z}^n}$ satisfies Property 3.1. In particular, we get a primal solution method for the above mentioned quotient functions. Then, it suffices to determine negative alternating cycles in G with respect to arc-weights of the form $c(e) - f(M) \cdot d(e)$, if $e \not\in M$, resp. $-c(e) + f(M) \cdot d(e)$, if $e \in M$.

Secondly, let $G = G(E \cup E', A, E, E')$ denote the bipartite digraph corresponding to polymatroid intersections (cf. Fig. 1). The results of Section 2 imply that the minimization of a function on (general) independent flows and polymatroidal flows can be reduced to a minimization problem on independent flows in G. Thus, let P denote the set of all feasible, independent flows in G which have maximum flow value, i.e., $P = G(\alpha)$. We will show validity of the proposed general strategy under a mild additional assumption. If $f(P)$ is finite, then finiteness is a consequence of validity; otherwise finiteness is not known. We assume that the domain D of the objective function f contains all vectors $x \in R_+^A$ with $x \leq c$.

For $x \in P$ the auxiliary digraph G_x is discussed at the end of Section 2. We define

$$Q(x) := \{x \oplus \Delta x \mid \Delta x \ F\text{-}B\text{-feasible circuit flow in } G_x\}.$$

Theorem 2.7 and the subsequent remarks imply

$$P \subseteq x + \text{cone}_Z \ (Q(x) - x).$$

Unfortunately, $Q(x)$ is not necessarily a subset of P. Therefore a direct application of the general strategy may lead into troubles. We may detect $y \in Q(x)$ with $f(y) < f(x)$ but $y \not\in P$. Therefore, we have to show that, if x is not locally optimal (in $Q(x)$), then there exists $y' \in Q(x)$ with $f(y') < f(x)$ and $y' \in P$.

Let $\tilde{C}_x$ denote the set of all circuits in G_x on which an F-B-feasible circuit flow exists. For $C \in \tilde{C}_x$ a circuit flow on C with value $\lambda \in R_+$ on all arcs of C is denoted by $\Delta x(C, \lambda)$. We call $C \in \tilde{C}_x$ *negative* if

$$f(x) > f(x \oplus \Delta x(C, \mu))$$

for some $0 < \mu \leq \delta(C)$ where

$$0 < \delta(C) := \min\{\tilde{c}(a) \mid a \in C\}$$

is the minimum of the arc capacities $\tilde{c}$ (cf. Section 2) of the arcs of C. Unfortunately, negativity of C is dependent on μ, in general. There, we assume that f satisfies the reverse implication in (3.4), too, i.e.,

$$[\exists\, 0 < \mu : f(x) \leq f(x + \mu(y - x))] \Rightarrow f(x) \leq f(y) \tag{3.6}$$

for all $x, y \in D$. If R is a field then (3.6) is implied by (3.4). For linear objective functions it suffices to assume the validity of the cancellation rule ($\mu\alpha = \mu\beta \Rightarrow \alpha = \beta$) for all $\alpha, \beta, \mu \in R$ with $0 < \mu$. Then

$$f(x) < f(x \oplus \Delta x(C, \lambda))$$

for all $0 < \lambda$ provided that C is a negative circuit.

Clearly, if G_x contains a negative circuit $C \in \tilde{C}_x$, then it contains a negative circuit $C' \in \tilde{C}_x$ of shortest length. For a proof of the following theorem we need the validity of a strengthened form of Property 3.1.

Property 3.1′. *Let $x \in D$ and $S \subsetneq D$. If $f(x) \leq f(y)$ for all $y \in S$, then $f(x) \leq f(z)$ for all $z \in D$ with $r \cdot z \in x + \operatorname{cone}_R(S - x)$ for some $r \in \mathbb{Z}_+$.*

If R is a field then Property 3.1′ is implied by Property 3.1.

Theorem 3.3. *Let f satisfy Property 3.1′ and (3.6). Let $x \in P$ ($= G(\alpha)$, α maximum) and let $C \in \tilde{C}_x$ be a negative circuit of shortest length. Then $\Delta x(C, \lambda)$ is admissible ($\lambda \in R_{++}$).*

Proof. Admissibility is a property of the considered circuit C. We use the denotations introduced in the definition of admissibility (cf. Theorem 2.8) with respect to G_x. We remark that C satisfies (2.17). We prove that G_1 does not contain a directed circuit. For G_2 an analogue argument can be given.

We suppose that G_1 contains a circuit F. Each arc $[(i, j), (k, r)]$ of that circuit F corresponds to an arc $(i, r) \in D_1$ (wavy arcs in the example in Fig. 6). For each such arc we find a subpath of C completing a new directed circuit. These

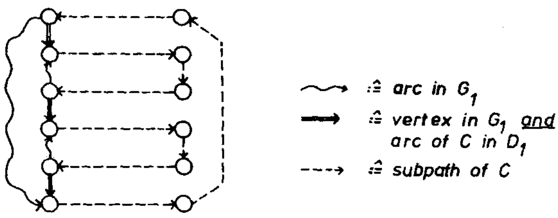


Fig. 6. Simple example for a circuit F in G_1. Each forward and backward arc in C is covered *once* by the new circuits.

directed circuits $C_1, C_2, \ldots, C_k \in \tilde{C}_x$ cover each forward and backward arc in C the same number of times, say r times. Therefore

$$x \oplus \Delta x(C, \mu \cdot r) = x \oplus \left[\sum_{i=1}^{k} \Delta x(C_i, \mu)\right]$$

for all $\mu \in R_+$. Obviously, the length of each C_i is strictly smaller than the length of C. Therefore, none of the circuits C_i is negative. Thus, Property 3.1′ and (3.6) lead to the contradiction that C is not negative. □

Theorem 2.3 shows that, if x is not locally optimal then we can construct a negative circuit which yields admissible circuit flows. Once a negative circuit is found this construction proceeds by using (shortcutting) arcs in D_1 or D_2 in order to find negative circuits of less length. Then, we achieve feasibility by choosing $0 < \lambda \leq \delta(C)$. Now Theorems 2.8 and 3.3 imply the following result.

Theorem 3.4. *Let f satisfy* (3.6) *and Property* 3.1′. *Then* $x \in G(\alpha)$ *is a minimum solution iff* $\tilde{C}_x$ *does not contain a negative circuit.*

Theorem 3.4 provides a local characterization of global optimality. Therefore, the following variant of the general strategy is valid.

Negative Circuit Method 3.5.

Step 1. Find $x \in G(\alpha)$.

Step 2. If $\tilde{C}_x$ does not contain a negative circuit, stop.

Step 3. Otherwise find a negative circuit $C \in \tilde{C}_x$ and its minimum arc capacity $\delta(C)$; $x := x \oplus (\Delta x(C, \delta))$ and go to Step 2.

We observe that the Negative Circuit Method is finite if $R = \mathbb{Z}$. In the general case, finiteness is not assured. We remark that a similar statement was already necessary in [7] for the special case of a linear function $f : R^n \to R$. Fujishige

[7] develops two solution methods: an augmenting path method and a Negative Circuit Method. The latter method coincides with the Negative Circuit Method 3.5 if f is linear. For the case of uncapacitated polymatroid intersections, $R = \mathbb{Z}$ and linear objective functions $f : \mathbb{Z}^n \to \mathbb{R}$ the minimization problem is previously solved in Schönsleben [17], too. He provides a nonpolynomial complexity bound (in the denotation of Section 2: $O(n^3\kappa(Z + n^2))$).

Recently, Frank [3] developed an elegant method using vertex-potentials in the case of linear objective functions $f : \mathbb{Z}^n \to \mathbb{Z}$ but for an even more general combinatorial structure introduced by Edmonds and Giles [2]. The method is quasi-polynomial and, if $0 \leqslant x \leqslant 1$, polynomial.

4. Concluding remark

After submitting this paper, we learnt about the papers of Lawler and Martel [15] as well as Frank [3]. Frank's discussion of the weighted combinatorial optimization problem of Edmonds and Giles [2] leads us to the question whether an approach, as given here, can be extended to that structure. The positive answer including a negative circuit method will be discussed in a forthcoming paper.

Acknowledgment

We thank two anonymous referees for several helpful comments and proposals. In particular, the proof of Theorem 2.1 looks now much simpler than in a preliminary version of our paper.

References

[1] J. Edmonds, Submodular functions, matroids and certain polyhedra, Proc. Calgary Internat. Conf. on Combinatorial Structures and their Applications (Gordon and Breach, New York, 1970) pp. 66–87.

[2] J. Edmonds and R. Giles, A min–max relation for submodular functions on graphs, Ann. Discrete Math. 1 (1977) 185–204.

[3] A. Frank, An algorithm for submodular functions on graphs, Ann. Discrete Math. 16 (1982) 97–120.

[4] H. Friesdorf and H. Hamacher, Weighted min cost flows, Rept. 81-7, Mathematisches Institut, Universität zu Köln (1981).

[5] S. Fujishige, A primal approach to the independent assignment problem, J. Oper. Res. Soc. Japan 20 (1977) 1–14.

[6] S. Fujishige, An algorithm for finding an optimal independent linkage, J. Oper. Res. Soc. Japan 20 (1977) 59–75.

[7] S. Fujishige, Algorithms for solving the independent-flow problem, J. Oper. Res. Soc. Japan 21 (1978) 189–203.

[8] R. Giles, Submodular functions, graphs and integer polyhedra, Ph.D. thesis, University of Waterloo, Waterloo, Ontario, 1975.

[9] H. Hamacher, Flows in regular matroids, Mathematical Systems in Economics (Hain, Meisenheim am Glan, 1981).

[10] R. Hassin, On network flows, Ph.D. thesis, Yale University, 1978.

[11] M. Iri, A practical algorithm for the Menger-type generalization of the independent assignment problem, Math. Programming Stud. 8 (1978) 88–105.

[12] M. Iri and N. Tomizawa, An algorithm for finding an optimal independent assignment, J. Oper. Res. Soc. Japan 19 (1976) 32–57.

[13] E.L. Lawler, Combinatorial Optimization: Networks and Matroids (Holt, Rinehart and Winston, New York, 1976).

[14] E.L. Lawler, An introduction to polymatroidal flows, Rept. BW 129/80, Department of Operations Research, Mathematisch Centrum, Amsterdam (1980).

[15] E.L. Lawler and C.U. Martel, Computing maximal 'polymatroidal' network flows, Memorandum No. UCB/ERL M 80/52, Electronics Research Laboratory, University of California, Berkeley, CA 94720 (1980).

[16] C.U. Martel, Generalized network flows with an application to multiprocessor scheduling, Ph.D. thesis, University of California at Berkeley, 1980.

[17] P. Schönsleben, Ganzzahlige Polymatroid-Intersektions Algorithmen, Ph.D. thesis, Eidgenössische Technische Hochschule Zürich, Zürich, 1980.

[18] U. Zimmermann, Linear and Combinatorial Optimization in Ordered Algebraic Structures, Annals of Discrete Mathematics 10 (North-Holland, Amsterdam, 1981).

This Page Intentionally Left Blank

Annals of Discrete Mathematics
Previous Volumes in this Series

Vol. 12: Theory and Practice of Combinatorics
edited by A. ROSA, G. SABIDUSI *and* J. TURGEON
1982 x + 265 in preparation

Vol. 13: Graph Theory
edited by B. BOLLOBÁS
1982 in preparation

Vol. 14: Combinatorial and Geometric Structures and their Applications
edited by A. BARLOTTI
1982 viii + 292 pages in preparation

Vol. 15: Algebraic and Geometric Combinatorics
edited by E. MENDELSOHN
1982 xiv + 378 in preparation