AN APPROACH TO OPTIMAL PARTITIONING OF HYPERGRAPHS

Giuseppe Alia and Piero Maestrini

Istituto di Elaborazione della Informazione
del CNR, Pisa, Italy

Key words and phrases: hypergraphs, network flow, cutsets, computer-aided design.

Summary.

The problem of determining optimal partitions of hypergraphs (or, more simply of ordinary graphs), is relevant in several areas, such as computer aided design of printed boards, information retrieval and program paging. In many cases there exist optimal or near optimal partitions, subject to the constraint that each block is an LS set. Intuitively, an LS set is a subset of nodes of the given hypergraph, more strongly connected to each other that to the nodes in the complementary subset.

This paper presents a polynomial-bounded procedure to determine all the LS sets in a given hypergraph.

The proposed procedure, that is considerably faster than those previously known, is based upon the representation of the given hypergraph with a "flow equivalent tree", into which a collection of subsets including all the LS sets is easily identified.

1. Introduction.

Many problems arising in different fields of computer science are amenable to finding a partition of the nodes of a graph or, more generally, of an hypergraph [1]. For example, in computer aided design of digital circuits it is often required to find a partition of a given circuit into subcircuits, such that the number of wires, or printed tracks, connecting the different subcircuits is as low as possible [2]. The circuit consists of a number of multiterminal integrated packages, and of various "signals" (wires). Each signal interconnects two or more modules. The problem is conveniently modeled by an hypergraph, where the nodes represent the modules and the arcs represent the signals. If n different signals interconnect the same modules, an unique arc weighted with n is drawn in the hypergraph. A similar model, using ordinary graphs, is useful when paging programs with the constraint that the number of jumps from any page to a different page is kept to a minimum [3], and in information storage and retrieval, whenever it is required to partition a set of documents into homogeneous subcollections, such that references to documents of different subcollections are as few as possible [4].

Consider an hypergraph $G = \{N,A\}$, where N is the set of the nodes and A is the set of the arcs, and each arc is incident to two or more nodes in N. Assume that the arcs in A are weighted with positive integers. For a given subset $S \subseteq N$, $w(S)$, called the weight of S, is defined as the sum of the weights of the arcs incident to at least a node in S and to at least a node in N-S. A subset $S \subseteq N$ is said to be an L S-set if the inequality $w(T) > w(S)$ holds for all proper subsets $T \subset S$. Intuitively, LS sets are subsets of nodes which are more strongly connected to each other, than to nodes in the complementary subset. LS sets were first introduced by Luccio and Sami in a paper [5] dealing with partitioning of networks. The importance of LS sets has been underlined by Lawler [6], who has shown that in many cases, where it is required to find an optimal partition of the nodes of an hypergraph, there exist optimal or near optimal solutions where each block of the partition is an LS set.

Lawler [6] has also found a polynomial-bounded algorithm to determine all the LS sets in a given hypergraph. This paper presents a more efficient algorithm to solve the same problem, through an approach based upon finding an equivalent representation of the hypergraph, called a cut-tree.

2. Cutsets and cut-trees.

Let $S \subset N$ be a subset of nodes in a given hypergraph and denote by $A(S)$ the set of arcs that are incident to one or more nodes in S. Let s and t be two distinct nodes in N, with $s \in S$, $t \in N-S$. Then the arcs in the subset $A(S) \cap A(N-S)$ are said to be a cutset disconnecting s and t, denoted by S,N-S. The sum of the weights of the arcs in S,N-S is called the capacity of the cutset, and is denoted by $c(S,N-S)$. It is immediately seen that $w(S)=w(N-S)=$ $=c(S,N-S)=c(N-S,S)$.

Among the cutsets disconnecting s and t, there exist one or more whose capacity is minimum. Such cutsets will be called, in the following, minimal cutsets disconnecting s and t. A procedure to determine a minimal cutset disconnecting s and t has been presented by Lawler [6]. In this procedure, the weighted hypergraph $G=\{N,A\}$ is transformed in a weighted directed graph $G^+=\{N^+,A^+\}$, with $N^+ \supseteq N$, $A^+ \supseteq A$, and the problem is reduced to a maximal flow computation in G^+. Taking s as the source and t as the sink, the labeling algorithm [7] yields the maximal flow value between s and t, coinciding with the capacity C_0^+ of minimal cutsets disconnecting s and t in G^+, and locates one of such cutsets, denoted S^+,N^+-S^+. As proved by Lawler, S,N-S, with $S=S^+ \cap N$ is a minimal cutset disconnecting s and t in G, and $C_0=C_0^+$ is the capacity of this cutset. In order to preserve the flow network analogy, C_0 will also be called the maximal flow value between s and t in G, denoted by $v(s,t)$. It should be noted that, although the flow computation is actually carried out in a directed flow network, interchanging the role of source and sink between s and t does not affect the maximal flow value.

Let $G=\{N,A\}$ be a weighted hypergraph and assume that the maximal flow between all pairs of nodes in N is computed: this defines a symmetrical function v, from N^2 to integers, called the flow function of G.

Two weighted hypergraphs, $G=\{N,A\}$ and $G'=\{N,A'\}$ are called flow equivalent if their flow functions coincide. Note that G' can possibly degenerate into an undirected weighted graph. It is easily seen that the procedure due to Gomory and Hu [8], to determine a flow equivalent tree of a given undirected flow network, can be extended to weighted hypergraphs. The discussion of the extended procedure [9] is omitted for the sake of brevity. This procedure yields flow-equivalent trees in a special class, called cut-trees. The following properties of cut-trees are an immediate extension of similar results by Gomory and Hu:

Theorem 1. Let $G=\{N,A\}$ be a weighted hypergraph and $T=\{N,A'\}$ be a cut-tree of G. Then;

- the maximal flow between the arbitrary nodes x_i and x_j in G equals the smallest arc capacity in the unique path joining x_i and y_i in T;
- if X and N-X are the subsets of nodes generated by disconnecting, in T, an arbitrary arc a_{ij} of capacity v_{ij}, then X,N-X is a minimal cutset in G, of capacity v_{ij}.

3. Properties of LS-sets.

In the context of this paper, the reason of considering a cut-tree of a given weighted hypergraph consists in the fact, to be shown in this section, that any cut-tree retains much of the information required to identify the LS sets. In order to derive the properties leading to LS sets identification in a cut-tree, some preliminary results need to be established. The following lemma 1 is an immediate generalization to the case of hypergraphs of a fundamental result stated by Gomory and Hu [8] for ordinary undirected graphs.

Lemma 1. Let X,N-X be a minimal cutset disconnecting the nodes x and y in a weighted hypergraph, with $x \in X$, and let x_1 and x_2 be arbitrary nodes in X. Then there exists a minimal cutset X_1,$N-X_1$ disconnecting x_1 and x_2, with $x_1 \in X_1$, such that either $x_1 \subset X$ or $N-X_1 \subset X$.

Before proving the further result stated by the subsequent lemma 2, the following notations need to be defined. For arbitrary subsets $N_i \subset N$ and $A_i \subset A$, define $A(N_i')=A-A(N_i)$, denote by $w'(A_i)$ the sum of the weights of the arcs in A_i and let N_i^+ denote either the literal N_i or N_i'. Finally, if P,Q,S,T are subsets of nodes, define:

$$w'(P^+,Q^+,S^+,T^+)=w'(A(P^+)\cap A(Q^+)\cap A(S^+)\cap A(T^+))$$

For example, with the preceding notations, $w'(P,Q',S,T')=$
$=w'(A(P)\cap(A-A(Q))\cap A(S)\cap(A-A(T)))$ is the set of arcs incident to at least a node in P and S and not incident to any node in Q and T.

Lemma 2. Let $X \subset N$ be an LS set of weight $w(X)=w_0$ in $G=\{N,A\}$ and $Y,N-Y$ be a minimal cutset disconnecting the arbitrary nodes y_1 and y_2, with $y_1 \in X, y_2 \in N-X$, $y_1 \in Y$, $y_2 \in N-Y$. Then necessarily $c(Y,N-Y) \leq w_0$ and $Y \supseteq X$.

Proof. Since $X,N-X$ is a cut of capacity w_0 disconnecting y_1 and y_2, the inequality $v(y_1,y_2)=c(Y,N-Y) \leq w_0$ holds. As $w(Y)= =c(Y,N-Y)$, it follows that relation $Y \subset X$ cannot hold, since it contradicts the hypothesis that X is an LS set. Further, assume that relation $Y \supseteq X$ does not hold, and define subsets P,Q,S,T as follows:

$P \cup S=X_1$ $Q \cup T=N-X$, $P \cup Q=Y$, $S \cup T=N-Y$. Then:

$$
\begin{aligned}
&c(P \cup Q, S \cup T)-c(P \cup Q \cup S,T)=\\
&=w'(P,Q',S,T')+w'(P,Q',S',T)+w'(P',Q,S,T')+\\
&+w'(P',Q,S',T)+w'(P,Q,S,T')+w'(P,Q,S',T)+\\
&+w'(P,Q',S,T)+w'(P',Q,S,T)+w'(P,Q,S,T)-\\
&-w'(P,Q',S',T)-w'(P',Q,S',T)-w'(P',Q',S,T)-\\
&-w'(P,Q,S',T)-w'(P,Q',S,T)-w'(P',Q,S,T)-\\
&-w'(P,Q,S,T)= \qquad 1)\\
&=w'(P,Q',S,T')+w'(P',Q,S,T')+w'(P,Q,S,T')-\\
&-w'(P',Q',S,T).
\end{aligned}
$$

Since $X=P \cup S$ is an LS set, $w(P \cup S)<w(P)$ or equivalently;

$$
\begin{aligned}
&w(P \cup S)-w(P)=\\
&=w'(P,Q,S',T')+w'(P,Q',S',T)+w'(P',Q,S,T')+\\
&+w'(P',Q',S,T)+w'(P,Q,S,T')+w'(P,Q',S,T)+\\
&+w'(P,Q,S',T)+w'(P',Q,S,T)+w'(P,Q,S,T)-\\
&-w'(P,Q',S,T')-w'(P,Q,S',T')-w'(P,Q',S',T)-\\
&-w'(P,Q,S',T)-w'(P,Q,S,T')-w'(P,Q',S,T)-\\
&-w'(P,Q,S,T)= \qquad 2)\\
&=w'(P',Q,S,T')+w'(P',Q',S,T)+w'(P',Q,S,T)-\\
&-w'(P,Q',S,T')<0
\end{aligned}
$$

Replacing inequality 2) into expression 1) yields

$$c(P \cup Q, S \cup T)-c(P \cup Q \cup S,T) > 2w'(P',Q,S,T')+ w'(P,Q,S,T')+w'(P',Q,S,T) \geq 0$$

From the preceding inequality it is seen that $P \cup Q$, $S \cup T$ cannot be a minimal cutset disconnecting y_1 and y_2, and it is concluded that relation $Y \supseteq X$ necessarily holds.

In order to make the following analysis easier, the LS sets need to be classified in two categories.

Definition 1. An LS set of weight w_0 is called a *min-cut LS set* if there exists at least one pair $\{x,y\}$, with $x \in X, y \in N-X$, such that $v(x,y)=w_0$. Any other LS set is called a *non-min-cut LS set*. If X is a min-cut LS set of weight w_0, there exist one or more minimal cutsets of capacity w_0 disconnecting pairs of nodes $\{x,y\}$, with $x \in X$, $y \in N-X$. Since $c(X,N-X)=w_0$, $X,N-X$ is one of such minimal cutsets.

The following property holds for min-cut LS sets:

Theorem 2. If X is a min-cut LS set of weight w_0, then $v(x,y) \leq w_0$ for every pair $\{x,y\}$ with $x \in X$, $y \in N-X$, and $v(x_1,x_2) > w_0$ for every pair $\{x_1,x_2\}$ with $x_1 \in X, x_2 \in X$.

Proof. If $x \in X, y \in N-X$, the inequality $v(x,y) \leq w_0$ is immediate [7], since $X,N-X$ is a cutset of capacity w_0. In order to prove the second inequality, assume that for some $x_1 \in X, x_2 \in X$ the equality $v(x_1,x_2)=w_1 \leq w_0$ holds. As stated by lemma 1, there exists a minimal cutset $X_1,N-X_1$ with $x_1 \in X_1, x_2 \in N-X_1$, such that either relation $X_1 \subset X$ or $N-X_1 \subset X$ holds. The capacity of this cutset is $c(X_1,N-X_1)=w_1$. Since $w(X_1)=w(N-X_1)=w_1$

it is concluded that there exists a subset of X of weight $w_1 \leqslant w_0$ and the assumption that X is an LS set is contradicted.

Let $X \subset N$ be a min-cut LS set of weight w_0, and $A_i = \{a_{i1}, a_{i2}, \ldots a_{ip}\}$ be the subset of arcs of T, incident to a node in X and to a node in N-X. Then it follows from theorems 1 and 2 that, in the cut-tree, any arc in A_i has capacity smaller than, or equal to, w_0; furthermore disconnecting the arcs in A_i yields a forest, where the nodes in X belong to a subtree, and the arcs connecting nodes in X have capacity greater then w_0.

A similar property holds for the non-min-cut LS sets, as stated by the following theorem, whose proof is constructive and follows from lemma 2:

<u>Theorem 3</u>. Let $X \subset N$ be a non-min-cut LS set of weight w_0. Then there exist cutsets $Z_i, N-Z_i$ $(1 \leqslant i \leqslant m)$, with $c(Z_i, N-Z_i) < w_0$ for every i, such that $X = Z_1 \cap Z_2 \cap \ldots \cap Z_m$. Furthermore the inequality $v(x_1, x_2) > \max(c_i)$ holds for every pair $\{x_1, x_2\}$ with $x_1 \in X, x_2 \in X$.

The proof [9] is omitted for the sake of brevity.

Let $X \subset N$ be a non min-cut- LS set and $A_j = \{a_{j1}, a_{j2}, \ldots a_{jp}\} \subseteq A'$ be the subset of arcs of T, incident to a node in X and to a node in N-X. Also, let c_{jk} $(1 \leqslant k \leqslant p)$ be the capacity of arc a_{jk}. Then it follows from theorems 1 and 3 that disconnecting the arcs in A_j yields a forest, where the nodes in X belong to a subtree and the arcs connecting nodes in X have capacity greater than $\max(c_{jk})$. Moreover, X has weight $w_0 > \max(c_{jk})$.

4. <u>Identification of LS-sets</u>.

Lawler [6] has presented a procedure for determining all the LS sets in a given hypergraph G and he has shown that, if G has n nodes and m arcs, the complexity of this procedure is $O(m^2 n^4)$. The properties derived in the preceding section suggest an alternative approach to LS set determination, resulting in a shorter computation.

Given an hypergraph $G = \{N, A\}$ with n nodes and m arcs, a cut-tree of G, denoted by $T = \{N, A'\}$ needs to be constructed first. Recalling from Edmonds and Karp [10] and Lawler [6] that a maximal flow computation in G is $O(m^2 n^2)$ in length, and observing that determining a cut-tree requires n-1 maximal flow computations [8], each in a network not larger than G, it is concluded that the construction of T is $O(m^2 n^3)$ in length.

The identification of LS sets in the cut-tree is made possible by the property established by the following theorem, whose proof, immediately resulting from theorems 2 and 3, is omitted for the sake of brevity.

<u>Theorem 4.</u> Given in the cut-tree $T = \{N, A'\}$ a subtree $T_i = \{N_i, A'_i\}$, with $N_i \subseteq N$, $A'_i \subseteq A'$, such that, unless $T_i = T$, for some integer c_i:

- every arc $a_{ih} \in A'_i$ has capacity $c_{ih} > c_i$,
- every arc $a_p \in A' - A'_i$ incident to a node in N_i and to a node in $N - N_i$ has capacity $c_p \leqslant c_i$,

define $c_i^+ = \min(c_{ih})$ and let $\{T_{i1}, T_{i2}, \ldots, T_{ik}\}$, with $T_{ij} = \{N_{ij}, A'_{ij}\}$ and $N_{ij} \subseteq N_i, A'_{ij} \subseteq A'_i$ be the collection of subtrees such that for each $1 \leqslant j \leqslant k$:

- every arc $a_{ijh} \in A'_{ij}$ has capacity $c_{ijh} > c_i^+$

- every arc $a_q \in A'-A'_{ij}$ incident a node in N_{ij} and to a node $N-N_{ij}$ has capacity $c_q \leqslant c_i^+$.

Then if $X \subset N_i$ is an LS set, relation $X \subseteq N_{ij}$ necessarily holds for some $1 \leqslant j \leqslant k$.

As a preliminary operation to systematic application of theorem 4, a collection $\mathcal{N}$ of subsets of N is constructed as follows. First, define $\mathcal{N} = \{N\}$, assume $T_1 = T = \{N, A'\}$ and denote by c_1^+ the smallest capacity of arcs in A'. In the first iteration, the collection $\{T_{11}, T_{12}, \ldots, T_{1k_1}\}$ is determined, where $T_{1j} = \{N_{1j}, A'_{1j}\}$ is defined as in theorem 4, and the subsets $N_{11}, N_{12}, \ldots, N_{1k_1}$ are added to $\mathcal{N}$. In the second iteration, for each T_{1j} $(1 \leqslant j \leqslant k_1)$ the smallest capacity value of the arcs incident to a pair of nodes in T_{1j} is denoted by c_{1j}^+, the collection $\{T_{1j1}, T_{1j2}, \ldots, T_{1Jk_1}\}$ is determined and the subsets $N_{1j1}, N_{1j2}, \ldots, N_{1Jk_{1j}}$ are added to $\mathcal{N}$. The process is iterated, through successive splitting of subtrees, until degenerate subtrees consisting of a single node are obtained. Since there are at most n-1 distinct values of arc capacity in T, and each iteration scans n nodes, the computation required to construct the collection $\mathcal{N}$ is $0(n^2)$ in length.

It immediately follows from theorem 4 that the collection contains all the LS sets, although any one subset in $\mathcal{N}$ is not necessarily an LS set. In order to completely identify the LS sets, the weight of each subset in $\mathcal{N}$ needs to be evaluated by referring to the bypergraph G. Identification follows recalling that $N_i \in \mathcal{N}$ is an LS set if and only if $w(N_i) < w(N_j)$ for each $N_j \in \mathcal{N}$ such that $N_j \subset N_i$. Observe also that any such N_j derives from splitting of N_i in the construction of $\mathcal{N}$.

To evaluate in G the weight of any subset $N_i \in \mathcal{N}$, the arcs incident to every node in N_i are to be considered, and the nodes connected by such arcs are to be scanned, until a node in $N-N_i$ is found. Since G has n nodes and m arcs, and it is easily seen that the subsets in the collection $\mathcal{N}$ are at most 2n-1, this computation is $0(m\ n^3)$ in length.

It is concluded that the proposed algorithm to determine the LS sets, consisting of three cascaded stages, i.e. construction of a cut-tree, determination of the collection $\mathcal{N}$ and identification of LS sets in $\mathcal{N}$, is polynomial-bounded and the overall computation is $0(m^2 n^3)$ in length. This result compares favourably with the previously known results.

References.

[1] C. Berge, Graphes et Hypergraphes, Dunod, Paris (1970).

[2] G. Alia, G. Frosini and P. Maestrini, "Automated Module placement and wire routeing according to a structured biplanar scheme in printed boards", Computer Aided Design, Vol. 5, pp.152-159 July 1973.

[3] F. Luccio, Private communication.

[4] G. Martella and M.G. Sami "A proposal for the organization of Large Files in a Hierarchical Structure" Technical Report, Istituto di Elettrotecnica ed Elettronica del Politecnico di Milano, Italy Dec. 1971.

[5] F. Luccio and M.G. Sami "On the decomposition of Networks into Minimally interconnected Subnetworks", IEEE Trans. on Circuit Theory, Vol. CT16, pp. 184-188, May 1969.

[6] E.L. Lawler, "Cutsets and Partitions of Hypergraphs" Networks, Vol.3, pp. 275-285 1973.

[7] L.R. Ford and D.R. Fulkerson, Flow in

Networks, Princeton University Press Princeton (1962).

[8] R.E. Gomory and T.C. Hu, "Multi-terminal Network Flows", J. Soc. Indust. Appl. Math., Vol. 9, pp. 551-570, Dec. 1961.

[9] G. Alia and P. Maestrini, "On optimal partitions of weighted hypergraphs" Internal Report B 74-6, Istituto di Elaborazione dell'Informazione del CNR, Pisa, Italy, March 1974.

[10] J. Edmonds and R. Karp, "Theoretical improvements in Algorithmic Efficiency for Network Flow Problems", J. Assoc. for Computing Mach, Vol. 19, pp. 248-264 (April 1972).