KYBERNETIKA — VOLUME 15 (1979), NUMBER 6

A Characterization of Hypergraphs Generated by Arborescences

MACIEJ M. SYSŁO

The paper presents some properties and characterizations of hypergraphs which can be derived from arborescences. These hypergraphs were introduced to provide a method for comparison different algorithms of cluster analysis.

1. INTRODUCTION

Let $H = (X, \mathscr{E})$ denote a *hypergraph*, where X is a finite set of *vertices* and $\mathscr{E} = \{E_i : E_i \subseteq X,\ i \in I\}$ is a family of subsets of X such that $E_i \neq \emptyset$ and $\bigcup_{i \in I} E_i = X$. $\mathscr{E}$ is the set of *edges* of H.

An *arborescence* is a directed rooted tree such that all vertices can be reached by a directed path starting from a distinguished vertex called a *root*. Let outdeg(v) and indeg(v) denote the numbers of arcs going out of v and coming into v, resp. The set of vertices of an arborescence A can be partitioned into three classes: (i) one element class consisting of the root, (ii) the class of *inner* vertices of A, and (iii) the class of *terminal* vertices of A. Let the vertices of classes (i) and (ii) be called *nonterminal*. Notice, that v is nonterminal if and only if outdeg $(v) > 0$.

The reader is referred to [1] for other terms not defined here.

A special kind of hypergraphs has been introduced in [2] to define a distance between arborescences with the same set of terminal vertices. These hypergraphs were used to represent a grouping of objects by different cluster analysis methods and the distance between them has been introduced to compare different clusterings.

The main purpose of this note is to present some properties and characterizations of the hypergraphs generated by arborescences.

Let $X = \{x_1, x_2, \ldots, x_n\}$ be the set of terminal vertices of an arborescence and let $\mathscr{A}$ denote the class of all arborescences with X as the set of terminal vertices. An arborescence $A \in \mathscr{A}$ generates the hypergraph $H_A = (X, \mathscr{E}_A)$ as follows: each nonterminal vertex v of A generates outdeg$(v) - 1$ the same edges in $\mathscr{E}_A$. Such an edge consists of those elements of X which are the terminal vertices of the subarborescence generated by vertex v. The subarborescence generated by vertex v is obtained by treating v as a root.

Let $\mathscr{H}_A$ denote the class of all hypergraphs with X as the set of vertices which can be generated by arborescences in $\mathscr{A}$.

Let us notice that for a fixed arborescence, a vertex v of outdegree 1 generates no edge of the corresponding hypergraph, so that we can assume that $\mathscr{A}$ contains only minimal homeomorphs (i.e., arborescences without vertices of outdegree 1)

In what follows, let $A \in \mathscr{A}$ and let $H_A = (X, \mathscr{E}_A)$ be the hypergraph generated by A. It has been proved in [2] that

Property 2.1. $|\mathscr{E}_A| = n - 1$.

Let $\mathscr{E}(E)$ $(E \in \mathscr{E}_A)$ denote the subfamily of $\mathscr{E}_A$ consisting of E and all its subsets which belong to $\mathscr{E}_A$ and $\mathscr{F}(E)$ denote only maximal proper subsets of E which belong to $\mathscr{E}_A$.

One can easily prove

Property 2.2. $|\mathscr{E}(E)| = |E| - 1$ $(E \in \mathscr{E}_A)$.

Property 2.1 follows from Property 2.2 if $E = X$.

Example 2.1. Let $X_1 = \{1, 2, 3, 4, 5\}$ and $\mathscr{E}_1 = \{\{1, 2, 3, 4, 5\}, \{1, 2, 3, 4\}, \{1, 2\}, \{2, 3\}\}$. Hypergraph $H_1 = (X_1, \mathscr{E}_1)$ has Property 2.2 but it can not be derived from any arborescence with 5 terminal vertices.

Property 2.3. $|E| \geqq 2$ for each E in $\mathscr{E}_A$.

Property 2.4. There exists $E \in \mathscr{E}_A$ such that $E = X$.

The last two properties follow from the construction of H_A. They can be also easily derived from Property 2.2.

The next property of H_A follows from the construction and the obvious property of an arborescence.

Property 2.5. Each pair of edges E and F of H_A satisfies either $E \subseteq F$ or $F \subseteq E$ or $E \cap F = \emptyset$.

A collection of subsets of a given set which satisfy Property 2.5, i.e., a family of sets in which any pair of sets is either disjoint or one contains the other is called a collection of *nested sets*, see [3].

Example 2.2. Let $\mathscr{E}_2 = \{\{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3\}, \{4, 5\}\}$. Hypergraph $H_2 = (X_1, \mathscr{E}_2)$ has Property 2.5 but it can not be derived from any arborescence with 5 terminal vertices. Let us notice that H_2 satisfies also Properties 2.3 and 2.4. Fig. 2.1 shows the arborescence corresponding to $\mathscr{E}_2$ regarded as a collection of nested sets.

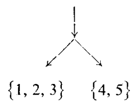

Fig. 2.1

Let $m(E)$ denote the multiplicity of E in $\mathscr{E}_A$. We shall prove now

Property 2.6. Every edge $E \in \mathscr{E}_A$ satisfies the equality

$$(1) \qquad 1 + m(E) = |\mathscr{F}(E)| + \left|E - \bigcup_{F \in \mathscr{F}(E)} F\right| .$$

Proof. It is easily seen that $|\mathscr{F}(E)|$ is the number of arcs of the arborescence A which go from vertex v generating edge E to other nonterminal vertices of A and $E - \bigcup_{F \in \mathscr{F}(E)} F$ is the set of terminal vertices adjacent from v. Thus, the value of the expression on the right hand side of (1) is equal to the number of edges outgoing from v so that the property is proved. □

Example 2.3. Let $\mathscr{E}_3 = \{\{1, 2, 3\}, \{1, 2\}, \{2, 4\}, \{3, 5\}\}$. Hypergraph $H_3 = (X_1, \mathscr{E}_3)$ has Properties 2.2 and 2.6 but has not Property 2.5.

Example 2.4. If $\mathscr{E}_4 = \{\{1, 2, 3, 4\}, \{1, 2, 3\}, \{1, 2\}, \{2, 4\}\}$ then $H_4 = (X_1, \mathscr{E}_4)$ has Properties 2.4 and 2.6 but has not 2.1 and 2.5.

Now we are ready to characterize hypergraphs in $\mathscr{H}_A$.

Theorem 2.1. Let $H = (X, \mathscr{E})$ be a hypergraph. $H \in \mathscr{H}_A$ if and only if H has Properties 2.2 and 2.5.

Proof. We have already shown that a hypergraph generated by an arborescence has Properties 2.2 and 2.5.

To prove the converse let us assume that $H = (X, \mathscr{E})$ has Properties 2.2 and 2.5 and we shall find the arborescence which generates H.

First, let us consider edges in $\mathscr{E}$ which are minimal, i.e., which contain no other edges. For each such an edge E we build the arborescence consisting of the vertex v_E corresponding to E and of arcs outgoing from v_E to vertices which belong to E. It follows from Property 2.2 that the vertices corresponding to such edges of H generate the hypergraph edges correctly.

The next step of the proof is by induction. Let $E \in \mathscr{E}$ and assume that all edges of H which are properly contained in E have been already considered. Property 2.5 guarantees that all edges of H can be considered in such a way. Moreover, on the base of the theorem assumptions we have

$$|\mathscr{E}(E)| = \sum_{F \in \mathscr{F}(E)} |\mathscr{E}(F)| + m(E) = \sum_{F \in \mathscr{F}(E)} |F| - |\mathscr{F}(E)| + m(E),$$

and hence

$$m(E) = |E| - \sum_{F \in \mathscr{F}(E)} |F| + |\mathscr{F}(E)| - 1 .$$

The term $|E| - \sum_{F \in \mathscr{F}(E)} |F|$ is equal to the number of elements of E which do not belong to any subset of E, an edge of H. Let us connect the vertex v_E which corresponds to E with vertices in $E - \bigcup_{F \in \mathscr{F}(E)} F$ and with subarborescences corresponding to subsets in $\mathscr{F}(E)$. We have outdeg $(v_E) = m(E) + 1$ and this proves the theorem. □

Another characterization can be obtained by applying the following lemma.

Lemma 2.1. Let $H = (X, \mathscr{E})$ be a hypergraph. If H has Properties 2.1, 2.4 and 2.6 then H has also Properties 2.2 and 2.5.

Proof. Let $E \in \mathscr{E}$ and E contain no other element of $\mathscr{E}$. Then, on the base of Property 2.6 we have $1 + m(E) = 0 + |E|$ and since $|\mathscr{E}(E)| = m(E)$ in this case, $|\mathscr{E}(E)| = |E| - 1$, so that E has Property 2.2. Then, let us apply an induction and suppose that E is minimal edge of $\mathscr{E}$ among the edges which have not been considered. Evidently, E contains some other edges of H which satisfy Property 2.2. Thus

$$|\mathscr{E}(E)| = \sum_{F \in \mathscr{F}(E)} |\mathscr{E}(F)| + m(E)$$

and on the base of Property 2.6 and by the inductive hypothesis we have

$$|\mathscr{E}(E)| = \sum_{F \in \mathscr{F}(E)} |F| - |\mathscr{F}(E)| + |\mathscr{F}(E)| + |E - \bigcup_{F \in \mathscr{F}(E)} F| - 1 =$$
$$= \sum_{F \in \mathscr{F}(E)} |F| + |E - \bigcup_{F \in \mathscr{F}(E)} F| - 1 = |E| - 1 + \sum_{F \in \mathscr{F}(E)} |F| - |\bigcup_{F \in \mathscr{F}(E)} F| .$$

Hence, if subsets in $\mathscr{F}(E)$ are not disjoint then $|\mathscr{E}(E)| = |E| - 1 + \alpha(E)$, where $\alpha(E) > 0$, and finally when we reach $E = X$ (it exists on the base of Property 2.4) we obtain $|\mathscr{E}_A| = |X| - 1 + \alpha$, where $\alpha > 0$. This contradicts the assumption that H has Property 2.1. Therefore, the set of edges $\mathscr{E}$ has Property 2.5 and also Property 2.2. □

As an immediate consequence of the above lemma we obtain

Theorem 2.2. Let $H = (X, \mathscr{E})$ be a hypergraph. $H \in \mathscr{H}_A$ if and only if H has Properties 2.1, 2.4 and 2.6.

One can easily show by a slight modification of the above proofs that hypergraphs in $\mathscr{H}_A$ can be also characterized as those having Properties 2.1, 2.5 and 2.6 or 2.4, 2.5 and 2.6.

3. EXTENSIONS

The results of the previous section can be easily extended if we replace arborescences by forests, i.e., the sets of arborescences. For instance, if a forest B has $p(B)$ components then the number of edges in the hypergraph generated by B is equal to the number of terminal vertices in all components minus $p(B)$.

(Received September 9, 1977.)

REFERENCES

[1] C. Berge: Graphs and hypergraphs. North-Holland, Amer. Elsevier, Amsterdam 1973.

[2] M. Karoński, Z. Palka: On Marczewski-Steinhaus type distance between hypergraphs. Zastosow. Matem. *16* (1977), 47—57.

[3] D. E. Knuth: The art of computer programming. Vol. 1: Fundamental algorithms, Reading (Mass.) 1968, 309.

Dr. Maciej M. Sysło, Institute of Computer Science, University of Wrocław, pl. Grunwaldzki 2/4, 50 384 Wrocław. Poland.