Archivum Mathematicum

Oldřich Sapák

Representation of the finite directed acyclic graph

Archivum Mathematicum, Vol. 13 (1977), No. 2, 111--116

Persistent URL: http://dml.cz/dmlcz/106965

ARCH. MATH. 2, SCRIPTA FAC. SCI. NAT. UJEP BRUNENSIS
XIII: 111—116, 1977

REPRESENTATION OF THE FINITE DIRECTED ACYCLIC GRAPH

OLDŘICH SAPÁK, Brno

(Received February 2, 1976)

A method is presented which allows to represent the finite directed acyclic graph by means of the real matrix. This representation facilitates easy to test a great number of binary relations on the graph and it is proved to be unique.

1. INTRODUCTION

The analysis of computer programs for the parallel processing leads many times to problems concerning finite directed acyclic graphs. As we shall deal with such graphs only, we shall always understand by the term graph a finite directed acyclic graph.

A graph having n nodes is usually represented in form of a $n \times n$ Boolean connectivity matrix C in this way: $c_{ij} = 1$ if and only if an arc (i, j) exists in this graph and $c_{ij} = 0$ in an opposite case. If we want to investigate some properties of a graph, it is better to use another representation of it.

2. CONSTRUCTION OF THE PROJECTION MATRIX

Let G be a graph with n nodes which we denote by $a_1, a_2, \ldots, a_n$. This defines an n-dimensional vector space V_n with the basis $a_1, \ldots, a_n$.

Definitions:

1. Let $v(a_i)$ denote the number of arcs beginning in the node a_i.

2. We say that the nodes a_i and a_j are in the relation R_0, if and only if there exists an arc (a_i, a_j) in the graph G.

3. For an arbitrary node a_i we define A_i to be the set of all immediate predecessors:

$$A_i = \{a_j \mid a_j R_0 a_i\}.$$

Construction: We shall construct the *flow* $T(a_i)$ in the node a_i in this way:

$$T(a_i) = \sum_{a_j \in A_i} \frac{T(a_j)}{v(a_j)} + a_i \qquad (*)$$

If there are no predecessors of a node a_k, then $A_k = \emptyset$ and $T(a_k) = a_k$. For an arbitrary node a_i of the graph G, the expression $T(a_i)$ is an element of V_n and it is of the form:

$T(a_i) = r_{i1}a_1 + r_{i2}a_2 + \dots + r_{in}a_n$, where r_{ik} are rational numbers.

We shall define the *projection* $P_i(j)$ of the node a_j in the direction a_i as a coordinate at a_i in the expression $T(a_j)$. It holds $r_{ik} = P_k(i)$.

The expression (*) can be transcribed by means of the projections in the following way:

$$P_i(j) = \sum_{a_k \in A_j} \frac{P_i(k)}{v(a_k)} + \delta_{ij},$$

where

$$\delta_{ij} = \begin{matrix} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j. \end{matrix}$$

Now let us construct the real projection matrix P.

P	$a_1 \dots a_i \dots a_n$
a_1	$\vdots$
$\vdots$	$\vdots$
a_j	$\dots\dots P_i(j)$
$\vdots$	
a_n	

This matrix P is the above-mentioned representation of the graph G.

3. PROPERTIES OF THE PROJECTION MATRIX

To demonstrate the reason for the construction of the projection matrix, we shall define several relations.

Definition: We say that

1. $a_iR_1a_j$, if there exists a path from the node a_i to the node a_j in the graph G;
2. $a_iR_2a_j$, if there exists no path from the node a_i to the node a_j in the graph G;
3. $a_iR_3a_j$, if all possible paths starting in the node a_i must reach the node a_j;
4. $a_iR_4a_j$, if there exists a path going from a_i to a_j, but there also exists another path going from a_i which passes by a_j, i.e. $a_iR_4a_j = (a_iR_1a_j)$ and $(a_i \text{ non } R_3a_j)$.

Theorem: *The following assertions hold for every pair of nodes a_i, a_j.*

a) $0 \leqq P_i(j) \leqq 1, \quad P_i(i) = 1,$

b) $a_i R_1 a_j \equiv P_i(j) > 0 \quad \text{and} \quad i \neq j,$

$a_i R_2 a_j \equiv P_i(j) = 0 \quad \text{or} \quad i = j,$

$a_i R_3 a_j \equiv P_i(j) = 1 \quad \text{and} \quad i \neq j,$

$a_i R_4 a_j \equiv 0 < P_i(j) < 1.$

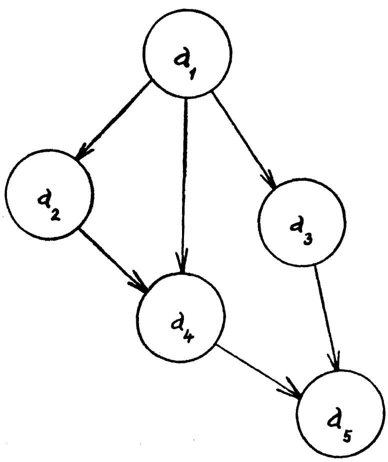
This theorem can be proved by means of the mathematical induction performed with respect to the number of arcs.

Remark: Projections can be also interpreted from the point of view of the theory of probability. Let us pass through the graph in the following way. In every node a_i having $v(a_i) = k$ it holds that the probability of the choice of each from the k following arcs is $1/k$, i.e. equal. In such case $P_i(j)$ represents the probability of the reaching of the node a_j under the condition that we have passed through the node a_i.

Remark: Some other relations can be also investigated by means of projections. In such a case the construction of $P_i(j)$ can be slightly modified.

The method has been originally designed for the testing of the relation R_5.

$a_i R_5 a_j = a_i R_4 a_j$ and $v(a_i) > 1$ and there exists no node a_k having this property: $a_i R_3 a_k$ and $a_k R_1 a_j$.



4. EXAMPLE

$$T(a_1) = a_1$$

$$T(a_2) = 1/3T(a_1) + a_2 = 1/3a_1 + a_2$$

$$T(a_3) = 1/3T(a_1) + a_3 = 1/3a_1 + a_3$$

$$T(a_4) = 1/3T(a_1) + T(a_2) + 1/2T(a_3) + a_4 = = 5/6a_1 + a_2 + 1/2a_3 + a_4$$

$$T(a_5) = 1/2T(a_3) + T(a_4) + a_5 = = a_1 + a_2 + a_3 + a_4 + a_5$$

$a_1 R_0 a_2$

$a_1 R_1 a_5$

$a_2 R_2 a_3$

$a_2 R_3 a_4$

$a_1 R_4 a_4$

P	a_1	a_2	a_3	a_4	a_5
a_1	1	0	0	0	0
a_2	1/3	1	0	0	0
a_3	1/3	0	1	0	0
a_4	5/6	1	1/2	1	0
a_5	1	1	1	1	1

5. EXISTENCE AND UNIQUENESS OF THIS REPRESENTATION

Existence is given by the construction of the matrix P. Before proving the uniqueness we shall introduce some definitions and lemmas.

1. For the graph G we define the $n \times n$ Boolean reachability matrix R in this way:

$$R_{ij} = 1 \equiv a_i R_1 a_j$$

$$R_{ik} = 0 \equiv a_i R_2 a_j$$

2. P is the projection matrix defined in the second section. It has the following "contravariant" connection to the matrices C and R: $R_{ij} = 1 \equiv a_i R_1 a_k \equiv P_i(j) = P_{ji} > 0$.

3. Let us call by *T-arc* the arc $a_i R_0 a_j$, if and only if there exists such a node a_k that it holds: $a_i R_1 a_k$ and $a_k R_1 a_j$.

4. For the graph G we define its *frame* G_1 as the graph containing all the arcs from G except T-arcs.

5. We shall understand by the transitive closure of G its frame G_1 to which all possible T-arcs are added.

Lemma 1: *Let G_1 and G'_1 be graphs without T-arcs with the same reachability matrix $R = R'$. Then it holds $G_1 = G'_1$.*

Proof: Suppose that G_1 and G'_1 are two different graphs without T-arcs and they have the same matrix R. Then there exist nodes a_i, a_j that $a_i R_0 a_j$ in G_1, but a_i non $R_0 a_j$ n G'_1 (or vice versa with respect to G_1 and G'_1). Since $a_i R_0 a_j$ in $G_1 \Rightarrow a_i R_1 a_i$ in $G_1 \Rightarrow R_{ij} = 1 \Rightarrow R'_{ij} = 1 \Rightarrow a_i R_1 a_j$ in G'_1. As a_i non $R_0 a_j$ in G'_1, there exists a node a_k, so that $a_i R_1 a_k$ and $a_k R_1 a_k$ in G'_1. Through matrix R the same relations hold in the graph G_1, where $a_i R_0 a_j$ also holds. Hence the arc $a_i R_0 a_j$ in G_1 is the T-arc, that yields a contradiction to the fact that G_1 has no T-arcs. Hence G_1 and G'_1 must be equal.

Lemma 2: *The reachability matrix R does not change if we add to or remove from a graph any T-arcs.*

The proof of this lemma is evident.

Corollary: *Let us consider the system $S(A)$ of all graphs with the same set of nodes A. The graphs in $S(A)$ with the same reachability matrix R will be put into the same class. The system of such classes forms a decomposition on the system $S(A)$. Each of these classes can be represented either by the matrix R or by the frame G_1 which is equal for all graphs of the given class, or by the transitive closure.*

Remark: Symbols referring to G_i are provided with a prime.

Lemma 3: *Let G and G' be two graphs with the same set of nodes and with the same projection matrices $P = P'$. Then G and G' have the same frame and for each node a_i it holds $v(a_i) = v'(a_i)$.*

Proof: To the matrix P we shall define a matrix R in the following way:

$$R_{ij} = 1 \equiv i \neq j \quad \text{and} \quad P_i(j) > 0,$$
$$R_{ij} = 0 \equiv i = j \quad \text{or} \quad P_i(j) = 0.$$

Matrix R defined in such a way is the reachability matrix of the both graphs G and G'. Thus the both graphs are in the same class of the decomposition; they have the same frame and they can differ only in T-arcs.

Suppose now that for some a_i it holds e.g. $v(a_i) < v'(a_i)$. At least one T-arc in the graph G' must go from the node a_i, so that the preceding inequality could hold. Since T-arcs go from the node a_i, there must also go some arc of the frame from the node a_i. Let it be the arc $a_i R_0 a_j$. Then it holds;

$P_i(j) = 1/v(a_i) > 1/v'(a_i) = P_i'(j)$, that is in the contradiction to $P = P'$ and hence for all a_i it holds $v(a_i) = v'(a_i)$.

Definition: To any arc of G we shall associate an integer which we shall call the *length* of the arc. Be $a_i R_0 a_j$ an arc. Then there exists a finite number of different (not necessary disjunct) paths leading from a_i to a_j in the graph G. To any of these paths we associate an integer – namely the number of arcs this path is composed of. We shall define the length of the arc $a_i R_0 a_j$ as the greatest of these integers.

Remark: All the arcs of the frame are of the length one; the T-arcs have the length greater or equal two.

Theorem: *Let G and G' be two finite directed acyclic graphs with the same set of nodes and with the same projection matrices $P = P'$. Then it holds $G = G'$.*

Proof: Suppose that $G \neq G'$. Then they can differ only in T-arcs. Let B be the set of all such T-arcs, that each of them appears just in one of the graphs G and G'.

$B = \{a_i R_0 a_j \mid (a_i R_0 a_j \text{ in } G \text{ and } a_i \text{ non } R_0 a_j \text{ in } G') \text{ or } (a_i \text{ non } R_0 a_j \text{ in } G \text{ and } a_i R_0 a_j \text{ in } G')\}$. Since $G \neq G'$, it holds $B \neq \emptyset$.

In B we shall find such a T-arc that all the other T-arcs from B are of the same or greater length (see Definition).

Let it be the arc $a_k R_0 a_m$ and let it be for instance an element of the graph G.

Be $C = \{a_i \mid a_k R_1 a_i \text{ and } a_i R_1 a_m\} \cup \{a_k, a_m\}$. Then for any pair of nodes a_p, $a_r \in C$ it holds $a_p R_0 a_r \notin B - \{a_k R_0 a_m\}$. [If $a_p R_0 a_r \in B - \{a_k R_0 a_m\}$, then the T-arc $a_p R_0 a_r$ would be of the smaller length then the arc $a_k R_0 a_m$.] From this fact and from the fact that $v(a_i) = v'(a_i)$ for all a_i, it follows that for all $a_p \in C - \{a_m\}$ it holds $P_k(p) = P_k'(p)$ and $P_k(m) = P_k'(m) + 1/v(a_k)$. Hence $P_k(m) > P_k'(m)$ which is a contradiction to the supposition $P = P'$. Hence it holds $G = G'$.

Corollary: *The finite directed acyclic graph is completly characterized by means of the projection matrix P which determines it as well as the connectivity matrix C.*

O. Sapák
662 95 Brno, Janáčkovo nám. 2a
Czechoslovakia