Časopis pro pěstování matematiky

Bohdan Zelinka
Caterpillars

Časopis pro pěstování matematiky, Vol. 102 (1977), No. 2, 179--185

Persistent URL: http://dml.cz/dmlcz/117956

Časopis pro pěstování matematiky, roč. 102 (1977), Praha

CATERPILLARS

BOHDAN ZELINKA, Liberec
(Received February 2, 1976)

A caterpillar is a tree C with the property that after deleting all terminal edges and all terminal vertices of C a snake (a tree consisting of one simple path) or the null-graph (a graph without vertices and without edges) is obtained. This concept was introduced by F. HARARY and A. J. SCHWENK in [1].

Evidently caterpillars together with the one-vertex graph form a class of trees which is closed under taking subtrees and under connected homomorphisms. Every star and every snake is a caterpillar.

If C is a caterpillar, then we denote by $B(C)$ the graph obtained from C by deleting all terminal vertices and all terminal edges. If $B(C)$ is the null-graph, then C is a tree with two vertices; this case is trivial. In the other cases $B(C)$ is a snake; we shall call it the body of C.

The vertices of $B(C)$ will be denoted by $v_0, v_1, \ldots, v_d$, where d is the length of $B(C)$ and the vertices v_i, v_{i+1} for $i = 0, 1, \ldots, d - 1$ are adjacent. If t_i is the number of terminal edges of C incident with v_i for $i = 0, 1, \ldots, d$, then C is uniquely determined by the vector $[t_0, t_1, \ldots, t_d]$. Note that t_0 and t_d must be different from zero; otherwise v_0 or v_d would be a terminal vertex of C and this would contradict the fact that v_0 and v_d belong to $B(C)$. Nevertheless, t_i for $1 \leqq i \leqq d - 1$ may be equal to zero. Evidently each $(d + 1)$-dimensional vector whose co-ordinates are non-negative integers and the first and the last of them are different from zero determines uniquely a caterpillar in which t_i have the described meaning. In general, two vectors may correspond to every caterpillar with at least three vertices; this depends on the choice of v_0 (for v_0 we choose one of the two terminal vertices of the body of C). The vector of a caterpillar does not depend on the choice of v_0, if and only if there exists an automorphism of C whose restriction onto the body of C is not an identity mapping. If we want to assign a unique vector to every caterpillar, we may take that one which precedes the other in the lexicographical ordering of the set of all $(d + 1)$-dimensional vectors. However, in the sequel, when we speak about the vector of a caterpillar, we mean anyone of the two vectors which are assigned to that caterpillar. A caterpillar with at least three vertices is a snake, if and only if its vector $[t_0, t_1, \ldots, t_d]$ has the property that $t_0 = t_d = 1$, $t_i = 0$ for $i = 1, \ldots, d - 1$. A caterpillar is a star, if and only if its vector is one-dimensional.

There exist various ways how to determine a tree. We shall mention some of them and show characterizations of caterpillars in terms of them.

L. NEBESKÝ [3] has defined tree algebras. A tree algebra (M, P) is an algebra with an element set M and with a ternary operation P which satisfies the following axioms:

I. $P(u, u, v) = u$;

II. $P(u, v, w) = P(v, u, w) = P(u, w, v)$;

III. $P(P(u, v, w), v, x) = P(u, v, P(w, v, x))$;

IV. $P(u, v, x) \neq P(v, w, x) \neq P(u, w, x) \Rightarrow P(u, v, x) = P(u, w, x)$.

Every finite tree T determines uniquely a tree algebra (M, P), whose element set is the vertex set of T and in which $P(u, v, w)$ is the common vertex of the path connecting u and v, the path connecting u and w and the the path connecting v and w in T. Conversely, every finite tree algebra determines uniquely a tree. Thus there is a one-to-one correspondence between finite trees and finite tree algebras. This was proved by L. Nebeský.

Theorem 1. *Let T be a finite tree with at least three vertices, let (M, P) be the tree algebra corresponding to T. The tree T is a caterpillar, if and only if for any nine elements $x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3$, where $x_1 \neq x_2 \neq x_3 \neq x_1$, $y_1 \neq y_2 \neq y_3 \neq y_1$, $z_1 \neq z_2 \neq z_3 \neq z_1$, the vertex $P(P(x_1, x_2, x_3), P(y_1, y_2, y_3), P(z_1, z_2, z_3))$ coincides with some of the vertices $P(x_1, x_2, x_3)$, $P(y_1, y_2, y_3)$, $P(z_1, z_2, z_3)$.*

Proof. Let T be a caterpillar. As $x_1 \neq x_2 \neq x_3 \neq x_1$, $P(x_1, x_2, x_3)$ cannot be a terminal vertex of T, because a terminal vertex of T can be contained only in such a path whose terminal vertex it is. Thus $P(x_1, x_2, x_3)$ and analogously also $P(y_1, y_2, y_3)$ and $P(z_1, z_2, z_3)$ belong to the body of T. The body of T is a snake, therefore for any three of its vertices there exists one of them which lies between the other two. This implies that at least one of the vertices $P(x_1, x_2, x_3)$, $P(y_1, y_2, y_3)$, $P(z_1, z_2, z_3)$ lies on the path connecting the other two and thus it is equal to $P(P(x_1, x_2, x_3), P(y_1, y_2, y_3), P(z_1, z_2, z_3))$. If T is not a caterpillar, then it contains a subtree isomorphic to the tree in Fig. 1; this was mentioned in [1]. If $x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3$ are such as is denoted in Fig. 1, then $P(x_1, x_2, x_3) = x_2$, $P(y_1, y_2, y_3) = y_2$, $P(z_1, z_2, z_3) = z_2$, but $P(x_2, y_2, z_2) = x_3$, which is different from x_2, y_2, z_2.

Another way of determining trees was described by E. A. SMOLENSKIÎ [4]. If $u_1, \ldots, u_n$ are terminal vertices of a tree T and d_{ij} is the distance between u_i nad u_j in T for $1 \leqq i \leqq n$, $1 \leqq j \leqq n$, then the matrix $\|d_{ij}\|$ is called the distance matrix of T. The tree T is uniquely (up to an isomorphism) determined by its distance matrix. In the following theorem the letter u with subscripts has this meaning and the letter v with subscripts has the meaning as in the definition of the vector of a caterpillar.

Theorem 2. *Let T be a tree with n terminal vertices, let $\mathbf{D} = \|d_{ij}\|$ be its distance matrix. The tree T is a caterpillar, if and only if any three pairwise distinct numbers i, j, k from the numbers $1, \ldots, n$ satisfy*

$$\min(d_{ij} + d_{jk} - d_{ik},\ d_{ij} + d_{ik} - d_{jk},\ d_{ik} + d_{jk} - d_{ij}) = 2 . \tag{1}$$

Proof. Let T be a caterpillar with the vector $[t_0, t_1, \ldots, t_d]$. Let $v_{l(i)}, v_{l(j)}, v_{l(k)}$ be the vertices of the body of T which are adjacent to u_i, u_j, u_k respectively. Without loss of generality let $l(i) \leqq l(j) \leqq l(k)$. Then

$$d_{ij} = 2 + l(j) - l(i)\,, \quad d_{jk} = 2 + l(k) - l(j)\,, \quad d_{ik} = 2 + l(k) - l(i)$$

and therefore

$$d_{ij} + d_{jk} - d_{ik} = 2\,,$$
$$d_{ij} + d_{ik} - d_{jk} = 2 + 2\,l(j) - 2\,l(i) \geqq 2\,,$$
$$d_{ik} + d_{jk} - d_{ij} = 2 + 2\,l(k) - 2\,l(j) \geqq 2\,.$$

Thus the equality (1) holds. Now suppose that (1) holds and prove that T is a caterpillar. Let again u_i, u_j, u_k be three pairwise distinct terminal vertices of T. Let $v = P(u_i, u_j, u_k)$; this symbol is taken from the tree algebra defined above. Evidently

$$d_{ij} = d(u_i, v) + d(u_j, v)\,, \quad d_{jk} = d(u_j, v) + d(u_k, v)\,, \quad d_{ik} = d(u_i, v) + d(u_k, v)\,,$$

where $d(x, y)$ denotes the distance of vertices x and y in T. Then

$$d_{ij} + d_{jk} - d_{ik} = 2d(u_j, v)\,, \quad d_{ij} + d_{ik} - d_{jk} = 2d(u_i, v)\,,$$
$$d_{ik} + d_{jk} - d_{ij} = 2d(u_k, v)\,.$$

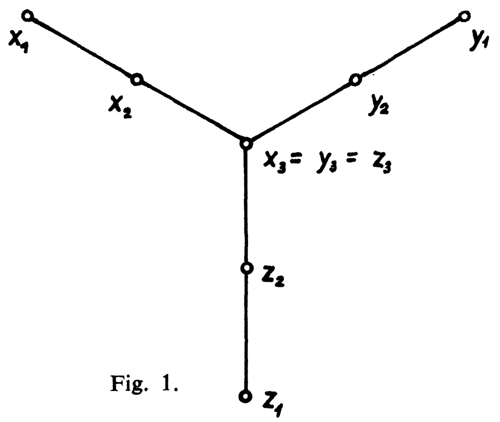

Fig. 1.

If (1) is fulfilled, then at least one of the vertices u_i, u_j, u_k has the distance 1 from $v = P(u_i, u_j, u_k)$. But this excludes the existence of a tree isomorphic to that in Fig. 1 and T must be a caterpillar.

Now we shall study some problems of embedding.

L. Nebeský has defined a completely separable tree as a tree which can be embedded into every block graph which has exactly two blocks and the same number of vertices as this tree. (A block graph is a graph, each of whose blocks is a clique. The problems of embedding trees into block graphs were studied in [5]. Here we shall define a stronger concept of a completely separable rooted tree.

A rooted graph is a graph in which one vertex is chosen and called the root of this graph. If this graph is a tree, it is called a rooted tree. A rooted tree is called completely separable, if it can be embedded into every rooted block graph which has exactly two blocks and the same number of vertices as this tree and its root is not a cut-vertex in it, in such a way the root of the tree and the root of the block graph coincide.

Theorem 3. *A rooted tree is completely separable, if and only if it is a caterpillar whose root is a terminal vertex adjacent to a terminal vertex of its body.*

Proof. Let C be a rooted caterpillar whose root is a terminal vertex adjacent to a terminal vertex of its body. Let the vector of C be $[t_0, t_1, \ldots, t_d]$, let the root of C be a vertex r adjacent to v_0. Let the number of vertices of C be n; then $d + 1 + \sum_{i=0}^{d} t_i = n$. For $j = 0, 1, \ldots, d$ let $n_j = j + 1 + \sum_{i=0}^{j} t_i$. Let G be a block graph having exactly two blocks, one with k vertices, the other with $n - k + 1$ vertices, where k is a positive integer, $2 \leqq k \leqq n - 1$. Let B_1 be the block of G with k vertices and B_2 the block of G with $n - k + 1$ vertices. Let a be a cut-vertex of G. Let the root r_0 of G be in B_1. If $k - 1 \leqq n_0$, then we identify the vertex v_0 of C with the vertex a of G, choose $k - 1$ vertices from the n_0 vertices of C adjacent with v_0, one of them being r, and identify them with the vertices of B_1 so that r is identified with r_0. The remaining vertices will be identified with vertices of B_2. The embedding is complete. If $k - 1 > n_0$, then there exists j such that $n_j < k - 1 \leqq n_{j+1}$. Then we identify v_{j+1} with a. From the t_{j+1} terminal vertices adjacent to v_{j+1} we choose $k - 1 - n_j$ ones. These vertices, the vertices $v_0, \ldots, v_j$ and all terminal vertices adjacent to some of the vertices $v_0, \ldots, v_j$ will be identified with the vertices of B_1, r being identified with r_0, and the remaining vertices will be identified with the vertices of B_2.

Now we shall prove that no other rooted tree is completely separable. A root of a completely separable rooted tree must be its terminal vertex. If we have a block graph with n vertices and two blocks, one of which has only two vertices and the root of this block graph is the vertex of the two-vertex block which is not a cut-vertex, then this root has the degree one. If we embed a tree with n vertices into this block graph, only a terminal vertex of this tree can be identified with this root. Let us have a rooted tree T which is not a caterpillar and suppose that it is completely separable. Then T contains a vertex w such that there exist at least three branches A_1, A_2, A_3 outgoing from w, each of which contains at least three vertices including w. The vertex w is not a root of T, because it is not terminal. Thus the root r of T belongs to a branch A_0 outgoing from w. The branch A_0 may coincide with some of the branches A_1, A_2, A_3; without loss of generality suppose $A_0 \neq A_1$, $A_0 \neq A_2$. Let A_0 have k vertices. Take a rooted block graph G with n vertices and two blocks, one of which has $k + 1$ vertices, contains the root r_0 of G and is denoted by B_1. No vertex of A_0 can be identified with the cut-vertex a of G; otherwise some vertices of B_1 would be identified with no vertex of T. Thus the whole A_0 is embedded into B_1 and a must be identified with a vertex a_0 of T adjacent to w and not belonging to A_0. Without loss of generality suppose that a_0 does not belong to A_2. Then A_2 must be embedded into the same block of G as B_0, but this is not possible, because this block has only $k + 1$ vertices and they are identified with the vertices of A_0 and with the vertex a_0. This is a contradiction with the assumption that T is completely separable. Now let C be a caterpillar with n vertices with the vector $[t_0, t_1, \ldots, t_d]$ and let its root be a terminal vertex adjacent to some v_j, where $j \neq 0$, $j \neq d$. Suppose that C is completely separable. Take a rooted block graph G with n vertices and two blocks, one of which has three vertices and contains the root of G. Then either v_{j-1}, or v_{j+1} must be identified with the cut-vertex of G. Now by a similar argument as in the preceding case we obtain a contradiction.

Now we shall consider embedding caterpillars into the graphs of n-dimensional cubes (or shortly n-cubes). A graph of the n-cube, where n is a positive integer, is the graph whose vertices are all n-dimensional vectors whose co-ordinates are zeros and ones and in which two vertices are joined by an edge if and only if they differ from each other in exactly one co-ordinate. Embedding trees into n-cubes was studied by I. Havel and P. Liebl [2]. Every finite tree is embeddable into an n-cube for some n. If T is a finite tree, then the minimal n such that T is embeddable into the n-cube will be called the dimension of T and denoted by dim T.

Theorem 4. *Let T be a tree with $k \geqq 2$ vertices. Then*

$$]\log_2 k[\leqq \dim T \leqq k - 1 .$$

These bounds cannot be improved.

Remark. The symbol $]x[$ denotes the least integer which is greater than or equal to x; some authors call it "the post-office function".

Proof. If l is a positive integer and $l <]\log_2 k[$, then $l < \log_2 k$. The number of vertices of the l-cube is $2^l < k$ and thus a graph with k vertices cannot be embedded into it. Therefore $]\log_2 k[\leqq \dim T$. The upper bound will be proved by induction. If $k = 2$, then T is isomorphic to the one-dimensional cube and $\dim T = 1$; thus the assertion is true. Now let $k > 2$. Let u be a terminal vertex of T, let e be the edge incident with u, let v be the other end vertex of e. By deleting u and e from T we obtain a tree T' with $k - 1$ vertices. Let $m = \dim T'$. According to the induction assumption, $m \leqq k - 2$. Consider a graph Q_{k-1} of the $(k-1)$-dimensional cube. Its vertices are $(k-1)$-dimensional vectors whose co-ordinates are zeros and ones. Let Q_{k-2} be the subgraph of Q_{k-1} induced by the set of all vertices of Q_{k-1} whose last co-ordinate is 0; this graph Q_{k-2} is a graph of the $(k-2)$-dimensional cube. We embed T' into Q_{k-2}. Let $[a_1, \ldots, a_{k-2}, 0]$ be the vertex of Q_{k-2} with which v is identified in this embedding (the co-ordinates $a_1, \ldots, a_{k-2}$ are zeros or ones). Then we identify u with $[a_1, \ldots, a_{k-2}, 1]$ and T is embedded into Q_{k-1}. Therefore $\dim T \leqq \dim T' + 1 \leqq k - 1$. The bounds cannot be improved, because a snake with k vertices can be embedded into the cube of the dimension $]\log_2 k[$ (as a part of its Hamiltonian path) and a star with k vertices cannot be embedded into the cube of the dimension smaller than $k - 1$ (in such a cube there exists no vertex of the degree at least $k - 1$).

Theorem 5. *Let k, m be positive integers such that $k \geqq 2$ and*

$$]\log_2 k[\leqq m \leqq k - 1 .$$

Then there exists a caterpillar C with k vertices such that $\dim C = m$.

Proof. For any positive integer h such that $2 \leqq h \leqq k - 2$ let $C(h)$ be a caterpillar with the vector $[t_0, t_1, \ldots, t_d]$, where $d = k - h - 1$, $t_0 = h - 1$, $t_d = 1$ and $t_i = 0$ for $i = 1, \ldots, d - 1$. The caterpillar $C(2)$ is a snake with k vertices and $\dim C(2) =]\log_2 k[$. The caterpillar $C(k-2)$ can be emebedded into the $(k-2)$-dimensional cube so that v_0 is identified with $[0, \ldots, 0]$, v_1 with $[1, 0, \ldots, 0]$, the terminal vertices adjacent to v_0 are identified with $[0, 1, 0, \ldots, 0]$, $[0, 0, 1, 0, \ldots, 0]$, … …, $[0, \ldots, 0, 1]$, the terminal vertex adjacent to v_1 is identified with $[1, 1, 0, \ldots, 0]$. But $C(k-2)$ cannot be embedded into the $(k-3)$-dimensional cube, because it contains the vertex v_0 of the degree $k - 2$. Therefore $\dim C(k-2) = k - 2$. Now let us take the caterpillars $C(h)$ and $C(h+1)$ for some h, $2 \leqq h \leqq k - 3$. The caterpillar $C(h+1)$ is obtained from $C(h)$ by deleting one terminal vertex and adding another. In the proof of Theorem 4 we have proved that by adding one terminal vertex the dimension of a tree increases at most by one; by deleting a vertex obviously it cannot increase. Thus $\dim C(h+1) \leqq 1 + \dim C(h)$. This implies that $\dim C(h)$ for $2 \leqq h \leqq k - 2$ attains all integral values in the interval $\langle]\log_2 k[, k - 2 \rangle$. There exists a caterpillar C with k vertices and $\dim C = k - 1$, namely a star. Thus the assertion is proved.

In the end we propose two problems.

Problem 1. A universal caterpillar for caterpillars with n vertices is a caterpillar into which each caterpillar with n vertices can be embedded. Determine the least number of vertices of a universal caterpillar for caterpillars with n vertices.

Problem 2. Characterize graphs whose spanning tree is a caterpillar. (This is a generalization of graphs with Hamiltonian paths.)

References

[1] *F. Harary* and *A. J. Schwenk:* The number of caterpillars. Discrete Math. *6* (1973), 359—365.

[2] *I. Havel* and *P. Liebl:* Embedding the polytomic tree into the *n*-cube. Časop. pěst. mat. *98* (1973), 307—314.

[3] *L. Nebeský:* Algebraic properties of trees. Acta Univ. Carolinae Philologica, Monographia XXV 1969.

[4] *Е. А. Смоленский:* Об одном способе линейной записи графов. Ж. вычисл. мат. и мат. физ. *3* (1962), 371—372.

[5] *B. Zelinka:* Embedding trees into block graphs. Czech. Math. J. *26* (1976), 273—279.

Author's address: 460 01 Liberec, Komenského 2 (Katedra matematiky VŠST).