Czechoslovak Mathematical Journal

Bohdan Zelinka

A distance between isomorphism classes of trees

Czechoslovak Mathematical Journal, Vol. 33 (1983), No. 1, 126–130

Persistent URL: http://dml.cz/dmlcz/101863

Czechoslovak Mathematical Journal, 33 (108) 1983, Praha

A DISTANCE BETWEEN ISOMORPHISM CLASSES OF TREES

BOHDAN ZELINKA, Liberec

(Received December 21, 1981)

In [1] a certain distance between isomorphism classes of graphs was introduced. Here we shall study an analogon of this distance for trees.

Consider the set $\mathscr{T}_n$ of all isomorphism classes of trees with n vertices, where $n \geqq 3$. For any two elements $\mathfrak{T}_1, \mathfrak{T}_2$ of $\mathscr{T}_n$ we introduce the number $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$ as the least integer with the property that there exists a tree with $n + \delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$ vertices which contains a subtree $T_1 \in \mathfrak{T}_1$ and subtree $T_2 \in \mathfrak{T}_2$. For the sake of simplicity we shall also use the notation $\delta_T(T_1, T_2)$ for two trees T_1 and T_2; this will mean $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$ for the classes $\mathfrak{T}_1, \mathfrak{T}_2$ such that $T_1 \in \mathfrak{T}_1$, $T_2 \in \mathfrak{T}_2$.

Theorem 1. *The functional δ_T is a metric on the set $\mathscr{T}_n$.*

Proof. Evidently $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) \geqq 0$ for any two elements $\mathfrak{T}_1, \mathfrak{T}_2$ of T_n and $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) = 0$ if and only if $\mathfrak{T}_1 = \mathfrak{T}_2$. Also evidently $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) = \delta_T(\mathfrak{T}_2, \mathfrak{T}_1)$. Now let $\mathfrak{T}_1, \mathfrak{T}_2, \mathfrak{T}_3$ be three elements of $\mathscr{T}_n$. There exists a tree T_{12} with $n + \delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$ vertices which contains a subtree $T_1 \in \mathfrak{T}_1$ and a subtree $T_2 \in \mathfrak{T}_2$ and there exists a tree T_{23} with $n + \delta_T(\mathfrak{T}_2, \mathfrak{T}_3)$ vertices which contains a subtree $T_2' \in \mathfrak{T}_2$ and a subtree $T_3 \in \mathfrak{T}_3$. The trees T_2, T_2' are isomorphic; take an isomorphic mapping of T_2 onto T_2' and identify each vertex of T_2 with its image in this mapping. We may suppose that T_{12} and T_{23} are vertex-disjoint. The graph T obtained in the described way from the trees T_{12} and T_{23} is evidently a tree. It has $n + \delta_T(\mathfrak{T}_1, \mathfrak{T}_2) + \delta_T(\mathfrak{T}_2, \mathfrak{T}_3)$ vertices and contains a subtree $T_1 \in \mathfrak{T}_1$ and a subtree $T_2 \in \mathfrak{T}_2$. Hence

$$\delta_T(\mathfrak{T}_1, \mathfrak{T}_3) \leqq \delta_T(\mathfrak{T}_1, \mathfrak{T}_2) + \delta_T(\mathfrak{T}_2, \mathfrak{T}_3)$$

and the triangle inequality holds.

Theorem 2. *Let $\mathfrak{T}_1 \in \mathscr{T}_n$, $\mathfrak{T}_2 \in \mathscr{T}_n$, $T_1 \in \mathfrak{T}_1$, $T_2 \in \mathfrak{T}_2$. Let k be a non-negative integer, $k < n$. Then the following two assertions are equivalent:*

(i) *There exists a tree T with $n + k$ vertices which contains a subtree isomorphic to T_1 and a subtree isomorphic to T_2.*

(ii) *There exists a tree T_0 with $n - k$ vertices such that both T_1 and T_2 contain subtrees isomorphic to T_0.*

Proof. (i) $\Rightarrow$ (ii). Let (i) hold. Let T_1', T_2' be subtrees of T isomorphic to T_1, T_2, respectively. As $k < n$, the trees T_1', T_2' have a non-empty intersection and this intersection is a subtree T_0' of T which has at least $n - k$ vertices. Choose a subtree T_0 of T_0' with exactly $n - k$ vertices. If we take an isomorphic mapping of T_1' onto T_1 and an isomorphic mapping of T_2' onto T_2, then the images of T_0 in these mappings are subtrees of T_1 and T_2 which are isomorphic to one another.

(ii) $\Rightarrow$ (i). Let (ii) hold. Without loss of generality suppose that T_1, T_2 are vertex-disjoint. Let T_0', T_0'' be subtrees of T_1, T_2, respectively, which are both isomorphic to T_0. Take an isomorphic mapping of T_0' onto T_0'' and identify each vertex of T_0' with its image in this mapping. The graph T obtained in this way is evidently a tree with $n + k$ vertices and it contains T_1, T_2 as subtrees.

Similarly as in [1] we may consider a graph $\mathscr{G}_n$ whose vertex set is $\mathscr{T}_n$ and in which two vertices $\mathfrak{T}_1, \mathfrak{T}_2$ are adjacent if and only if $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) = 1$.

Theorem 3. *The distance of any two vertices $\mathfrak{T}_1, \mathfrak{T}_2$ of $\mathscr{G}_n$ is equal to $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$.*

Proof. Let $\mathfrak{T}_1, \mathfrak{T}_2$ be two vertices of $\mathscr{G}_n$ and let $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) = k$. Then there exists a tree T with $n + k$ vertices which contains a subtree $T_1 \in \mathfrak{T}_1$ and a subtree $T_2 \in \mathfrak{T}_2$. In T exactly $n - k$ vertices are common to T_1 and T_2 (see Theorem 2). Further, there are k vertices of T_1 not belonging to T_2 and k vertices of T_2 not belonging to T_1. The vertices of T_1 not belonging to T_2 will be denoted by $u_1, \ldots, u_k$ in such a way that each u_i is adjacent either to a common vertex of T_1 and T_2, or to a vertex u_j with $j < i$; this can be easily done. The vertices of T_2 not belonging to T_2 will be denoted by $v_1, \ldots, v_k$ in such a way that each v_i is adjacent either to a common vertex of T_1 and T_2, or to a vertex v_j with $j > i$. Then for each $j = 1, \ldots, k$, the graph S_i obtained from T_2 by deleting the vertices u_i for $i \leqq j$ and adding the vertices v_i for $i \leqq j$ together with the edges joining them with each other and with the common vertices of T_1 and T_2 in T, is a tree. Evidently $S_k = T_1$, $\delta_T(T_2, S_1) = 1$, $\delta_T(S_i, S_{i+1}) = 1$ for $i = 1, \ldots, k - 1$. The vertices $T_2, S_1, \ldots, S_k = T_1$ (here we speak about trees as vertices of $\mathscr{G}_n$ instead of classes containing them; we do this for the sake of simplicity) form a path of the length k in $\mathscr{G}_n$ and thus the distance of $\mathfrak{T}_1$ and $\mathfrak{T}_2$ in $\mathscr{G}_n$ is at most $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$. Now suppose that the distance between $\mathfrak{T}_1$ and $\mathfrak{T}_2$ in $\mathscr{G}_n$ is l. There exists a path of the length l in $\mathscr{G}_n$ consisting of the vertices $T_1 = S_0', S_1', \ldots, S_k' = T_2$. We have $\delta_T(S_i', S_{i+1}') = 1$ for $i = 0, \ldots, k - 1$. Let S_i'' be a tree with $n + 1$ vertices which contains a subtree isomorphic to S_i' and a subtree isomorphic to S_{i+1}'. For each $i = 0, \ldots, k - 2$ we choose an isomorphism of the subtree of S_i'' isomorphic to S_{i+1}' onto the subtree of S_{i+1}'' isomorphic to S_{i+1}' and identify each vertex of the domain of this mapping with its image. Then we obtain a tree with $n + l$ vertices which contains a subtree from $\mathfrak{T}_1$ and a subtree from $\mathfrak{T}_2$. Thus the distance between $\mathfrak{T}_1$ and $\mathfrak{T}_2$ in $\mathscr{G}_n$ is at least $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$; together with the previous result this yields that this distance is equal to $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$.

A snake is a tree consisting of one path. Its length is the length of this path.

Theorem 4. *The diameter of the graph $\mathcal{G}_n$ is $n - 3$. There exists exactly one pair of vertices of $\mathcal{G}_n$ whose distance is $n - 3$.*

Proof. As $n \geqq 3$, each tree from $\mathcal{G}_n$ contains a subtree which is a snake of the length 2; it has three vertices. If $\mathfrak{T}_1 \in \mathcal{T}_n$, $\mathfrak{T}_2 \in \mathcal{T}_n$, $T_1 \in \mathfrak{T}_1$, $T_2 \in \mathfrak{T}_2$, then according to Theorem 2 there exists a tree with $2n - 3$ vertices which contains a subtree isomorphic to T_1 and a subtree isomorphic to T_2. Thus $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) \leqq n - 3$ for any two vertices $\mathfrak{T}_1$, $\mathfrak{T}_2$ of $\mathcal{G}_n$. Now let S_1 be the snake of the length $n - 1$ and let S_2 be a star with $n - 1$ edges. Any subtree of S_1 (or S_2) with more than three vertices is a snake (or a star, respectively) with more than two edges. Therefore (ii) from Theorem 2 for $k \leqq n - 3$ does not hold, thus (i) does not hold, either, and the isomorphism classes containing S_1 and S_2 have the distance exactly $n - 3$. Any tree with n vertices which is neither a snake nor a star contains a snake with four vertices and a star with four vertices as subtrees; hence the distance of its isomorphism class from any other isomorphism class from $\mathcal{T}_n$ is at most $n - 4$.

For every positive integer $k \geqq 3$ we shall define the tree $T(k)$. First we define the graph $T_0(k)$. The vertex set of $T_0(k)$ consists of all vectors of the dimensions $0, 1, \ldots$ $\ldots,]k/2[- 1$(the symbol $]x[$ denotes the least integer greater than or equal to x) whose coordinates are the numbers from the set $\{1, \ldots, k - 1\}$. Two vectors $\boldsymbol{u}, \boldsymbol{v}$ are adjacent if and only if one of them is obtained from the other by adding one coordinate. If k is odd, we take two disjoint copies of $T_0(k)$ and join the vertices corresponding to the zero vector in both of them. If k is even, we take a new vertex a and k pairwise disjoint copies of $T_0(k)$ and join a with the vertices corresponding to the zero vector in all of them. The tree thus obtained will be denoted by $T(k)$.

Lemma 1. *The tree $T(k)$ has the maximal number of vertices among all trees with the diameter at most k and the maximal degree at most k.*

Proof. Let T be a tree with the diameter k and the maximal degree k. If k is even, then T has one centre c and the distance of each vertex of T from c is at most $k/2$. As the maximal degree of T is k, for each $i = 1, \ldots, k/2$ there are at most $k(k - 1)^{i-1}$ vertices of T whose distance from c is i. Thus T has at most $1 + k \sum_{i=0}^{k/2-1} (k - 1)^{i-1}$ vertices and this is the number of vertices of $T(k)$. The proof for k odd is analogous.

By $\tau(k)$ we denote the number of vertices of $T(k)$ for each $k \geqq 3$. Evidently

$$\tau(k) = 1 + k \sum_{i=0}^{k/2-1} (k - 1)^{i-1} \quad \text{for } k \text{ even},$$

$$\tau(k) = 2 \sum_{i=0}^{k/2-1} (k - 1)^{i-1} \quad \text{for } k \text{ odd}.$$

Further, for $n \geqq 6$ we denote

$$\sigma(n) = \max \{k \in N \mid \tau(k) \leqq n\},$$

where N denotes the set of all positive integers.

Theorem 5. *Let ϱ be the radius of $\mathscr{G}_n$. Then*

$$\varrho \leqq n - \sigma(n) - 1\,.$$

Proof. Let $k = \sigma(n)$ and construct the tree C. If $\tau(k) = n$, then $C \cong T(k)$. If $\tau(k) < n$, then the tree C is an arbitrary tree with n vertices containing $T(k)$ as a subtree. Let T be an arbitrary tree with n vertices. If the diameter of T is greater than k, then both T and C contain a snake with $k + 1$ vertices as a subtree. If $\mathfrak{C}$ and $\mathfrak{T}$ aιe isomorphism classes containing C and T, respectively, then $\delta_T(\mathfrak{C}, \mathfrak{T}) \leqq \leqq n - k - 1$. If the diameter of T is less than k, then (as it has $n \geqq \tau(k)$ vertices) by Lemma 1 its maximal degree must be greater than k. Then both C and T contain a star with $k + 1$ vertices as a subtree and again $\delta_T(\mathfrak{C}, \mathfrak{T}) \leqq n - k - 1$. The distance of $\mathfrak{C}$ from the isomorphism class containing a snake and from one containing a star is evidently exactly $n - k - 1$. Thus the radius of $\mathscr{G}_n$ is at most $n - k - 1 = = n - \sigma(n) - 1$.

Conjecture 1. *The radius of $\mathscr{G}_n$ is equal to $n - \sigma(n) - 1$.*

In the sequel we shall study caterpillars. A caterpillar is a tree with the property that after deleting all of its terminal vertices (vertices of degree 1) a snake is obtained (a graph consisting of one vertex is also considered a snake). The snake just mentioned is called the *body of the caterpillar.*

Theorem 6. *Let $\mathfrak{T}_1 \in \mathfrak{T}_n$, $\mathfrak{T}_2 \in \mathfrak{T}_n$, $T_1 \in \mathfrak{T}_1$, $T_2 \in \mathfrak{T}_2$. Let T_1, T_2 be caterpillars and let $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) = k$. Then there exists a caterpillar T with $n + k$ vertices which contains a subtree isomorphic to T_1 and a subtree isomorphic to T_2.*

Proof. As $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) = k$, according to Theorem 2 there exists a tree T_0 with $n - k$ vertices such that both T_1 and T_2 contain subtrees isomorphic to T_0. We have $n - k \geqq 3$, therefore T_0 has at least two edges. As it is a subtree of a caterpillar, it is a caterpillar. Let $B(T_1)$, $B(T_2)$, $B(T_0)$ be the bodies of the caterpillars T_1, T_2, T_0, respectively. Let T be the tree constructed as in the proof of Theorem 2. If T is not a caterpillar, then there exists an edge e_1 of $B(T_1)$ not belonging to $B(T_2)$ and an edge e_2 of $B(T_2)$ not belonging to $B(T_1)$, such that they both are incident with a vertex v_0 of $B(T_0)$. Let v_1 (or v_2) be the end vertex of e_1 (or of e_2, respectively) distinct from v_0. By identifying the vertices v_1, v_2 in T a tree with $n + k - 1$ vertices is obtained which contains both T_1 and T_2 as subtrees; this is a contradiction with the assumption that $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2) = k$. Thus T is a caterpillar, which was to be proved.

Corollary. *The set of all isomorphism classes of caterpillars with n vertices induces a subgraph $\tilde{\mathscr{G}}_n$ of $\mathscr{G}_n$ with the property that the distance in $\tilde{\mathscr{G}}_n$ is the same as in $\mathscr{G}_n$. The diameter of $\tilde{\mathscr{G}}_n$ is $n - 3$.*

Now for every positive integer k we construct a caterpillar $\tilde{T}(k)$. The body of $\tilde{T}(k)$ is a snake of the length $k - 2$. The degree of any vertex of this body in $\tilde{T}(k)$ is k. Evidently the number of vertices of $\tilde{T}(k)$ is $k^2 - 2k + 3$.

Lemma 2. *The caterpillar $\tilde{T}(k)$ has the maximal number of vertices among all caterpillars with the diameter at most k and the maximal degree at most k.*

Proof. Evidently the diameter of a caterpillar is the length of its body plus two. This implies the assertion.

Theorem 7. *Let $\tilde{\varrho}$ be the radius of $\tilde{\mathscr{G}}_n$. Then*

$$\tilde{\varrho} \leqq n - \tilde{\sigma}(n) - 1 ,$$

where

$$\tilde{\sigma}(n) = \max \{k \in N \mid k^2 - 2k + 3 \leqq n\} .$$

Proof is analogous to that of Theorem 5.

Conjecture 2. *The radius of $\mathscr{G}_n$ is equal to $n - \tilde{\sigma}(n) - 1$.*

In the end we shall compare the distance δ_T with the distance introduced in [1] on the set of all isomorphism classes of undirected graphs with n vertices. The distance $\delta(\mathfrak{G}_1, \mathfrak{G}_2)$ of two such classes was defined as the least number k such that there exists a graph with $n + k$ vertices which contains an induced subgraph belonging to $\mathfrak{G}_1$ and an induced subgraph belonging to $\mathfrak{G}_2$.

Theorem 8. *For two elements $\mathfrak{T}_1, \mathfrak{T}_2$ of $\mathscr{T}_n$ for $n \geqq 7$ the distances $\delta_T(\mathfrak{T}_1, \mathfrak{T}_2)$, $\delta(\mathfrak{T}_1, \mathfrak{T}_2)$ are different in general.*

Proof. Let $\mathfrak{S}_1$ (or $\mathfrak{S}_1$) be the isomorphism class containing a snake (or a star, respectively) with n vertices. We know that $\delta_T(\mathfrak{S}_1, \mathfrak{S}_2) = n - 3$. Now let $S_1 \in \mathfrak{S}_1$, $S_2 \in \mathfrak{S}_2$. In S_1 take an independent set with the maximal number of elements; it has $]n/2[$ vertices. Identify each vertex of this set with one terminal vertex of S_2. We obtain a graph with $[3n/2]$ vertices which contains S_1 and S_2 as induced subgraphs. Thus

$$\delta(\mathfrak{S}_1, \mathfrak{S}_2) \leqq [3n/2] - n = [n/2] < n - 3 = \delta_T(\mathfrak{S}_1, \mathfrak{S}_2) .$$

Reference

[1] *Zelinka, B.:* On a certain distance between isomorphism classes of graphs. Časop. pěst. mat. *100* (1975), 371–373.

Author's addres: 460 01 Liberec, Felberova 2, ČSSR (katedra matematiky VŠST).