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

**Theorem 1.** The functional \( \delta_\mathcal{T} \) is a metric on the set \( \mathcal{T}_n \).

**Proof.** Evidently \( \delta_\mathcal{T}(\mathcal{T}_1, \mathcal{T}_2) \geq 0 \) for any two elements \( \mathcal{T}_1, \mathcal{T}_2 \) of \( \mathcal{T}_n \) and \( \delta_\mathcal{T}(\mathcal{T}_1, \mathcal{T}_2) = 0 \) if and only if \( \mathcal{T}_1 = \mathcal{T}_2 \). Also evidently \( \delta_\mathcal{T}(\mathcal{T}_1, \mathcal{T}_2) = \delta_\mathcal{T}(\mathcal{T}_2, \mathcal{T}_1) \). Now let \( \mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3 \) be three elements of \( \mathcal{T}_n \). There exists a tree \( T_{12} \) with \( n + \delta_\mathcal{T}(\mathcal{T}_1, \mathcal{T}_2) \) vertices which contains a subtree \( T_1 \in \mathcal{T}_1 \) and a subtree \( T_2 \in \mathcal{T}_2 \) and there exists a tree \( T_{23} \) with \( n + \delta_\mathcal{T}(\mathcal{T}_2, \mathcal{T}_3) \) vertices which contains a subtree \( T_2 \in \mathcal{T}_2 \) and a subtree \( T_3 \in \mathcal{T}_3 \). 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 \( n + \delta_\mathcal{T}(\mathcal{T}_1, \mathcal{T}_2) + \delta_\mathcal{T}(\mathcal{T}_2, \mathcal{T}_3) \) vertices and contains a subtree \( T_1 \in \mathcal{T}_1 \) and a subtree \( T_2 \in \mathcal{T}_2 \). Hence

\[
\delta_\mathcal{T}(\mathcal{T}_1, \mathcal{T}_3) \leq \delta_\mathcal{T}(\mathcal{T}_1, \mathcal{T}_2) + \delta_\mathcal{T}(\mathcal{T}_2, \mathcal{T}_3)
\]

and the triangle inequality holds.

**Theorem 2.** Let \( \mathcal{T}_1 \in \mathcal{T}_n, \mathcal{T}_2 \in \mathcal{T}_n, T_1 \in \mathcal{T}_1, T_2 \in \mathcal{T}_2 \). 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 \(G_n\) whose vertex set is \(\mathcal{F}_n\) and in which two vertices \(\mathcal{I}_1, \mathcal{I}_2\) are adjacent if and only if \(\delta_T(\mathcal{I}_1, \mathcal{I}_2) = 1\).

Theorem 3. The distance of any two vertices \(\mathcal{I}_1, \mathcal{I}_2\) of \(G_n\) is equal to \(\delta_T(\mathcal{I}_1, \mathcal{I}_2)\).

Proof. Let \(\mathcal{I}_1, \mathcal{I}_2\) be two vertices of \(G_n\) and let \(\delta_T(\mathcal{I}_1, \mathcal{I}_2) = k\). Then there exists a tree \(T\) with \(n + k\) vertices which contains a subtree \(T'_1 \in \mathcal{I}_1\) and a subtree \(T'_2 \in \mathcal{I}_2\). 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 \(T'_1\) 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 \(i = 1, \ldots, k\), the graph \(S_i\) obtained from \(T'_2\) by deleting the vertices \(u_i\) for \(i \leq j\) and adding the vertices \(v_i\) for \(i \leq j\) 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 \(G_n\) instead of classes containing them; we do this for the sake of simplicity) form a path of the length \(k\) in \(G_n\) and thus the distance of \(\mathcal{I}_1\) and \(\mathcal{I}_2\) in \(G_n\) is at most \(\delta_T(\mathcal{I}_1, \mathcal{I}_2)\). Now suppose that the distance between \(\mathcal{I}_1\) and \(\mathcal{I}_2\) in \(G_n\) is \(l\). There exists a path of the length \(l\) in \(G_n\) 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 \(S'_l\) 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 \(\mathcal{I}_1\) and a subtree from \(\mathcal{I}_2\). Thus the distance between \(\mathcal{I}_1\) and \(\mathcal{I}_2\) in \(G_n\) is at least \(\delta_T(\mathcal{I}_1, \mathcal{I}_2)\); together with the previous result this yields that this distance is equal to \(\delta_T(\mathcal{I}_1, \mathcal{I}_2)\).

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 $n \geq 3$, each tree from $\mathcal{G}_n$ contains a subtree which is a snake of the length 2; it has three vertices. If $\mathcal{T}_1 \in \mathcal{F}_n, \mathcal{T}_2 \in \mathcal{F}_n, T_1 \in \mathcal{T}_1, T_2 \in \mathcal{T}_2$, 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 $\delta_1(\mathcal{T}_1, \mathcal{T}_2) \leq n - 3$ for any two vertices $\mathcal{T}_1, \mathcal{T}_2$ 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 $k \leq n - 3$ 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 $\mathcal{F}_n$ is at most $n - 4$.

For every positive integer $k \geq 3$ 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 $0, 1, \ldots, \lfloor k/2 \rfloor - 1$ (the symbol $\lfloor x \rfloor$ denotes the least integer greater than or equal to $x$) whose coordinates are the numbers from the set $\{1, \ldots, k - 1\}$. Two vectors $u, v$ 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 $i = 1, \ldots, \lfloor k/2 \rfloor$ there are at most $k(k - 1)^{i-1}$ vertices of $T$ whose distance from $c$ is $i$. Thus $T$ has at most $1 + k \sum_{i=0}^{\lfloor k/2 \rfloor - 1} (k - 1)^{i-1}$ 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 $k \geq 3$. Evidently

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

Further, for $n \geq 6$ we denote

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

where $N$ denotes the set of all positive integers.
Theorem 5. Let \( \varrho \) be the radius of \( G_n \). Then
\[
\varrho \leq n - \sigma(n) - 1.
\]

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 \( \mathcal{C} \) and \( \mathcal{I} \) are isomorphism classes containing \( C \) and \( T \), respectively, then \( \delta_T(\mathcal{C}, \mathcal{I}) \leq n - k - 1 \). If the diameter of \( T \) is less than \( k \), then (as it has \( n \geq \tau(k) \) 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 \( \delta_T(\mathcal{C}, \mathcal{I}) \leq n - k - 1 \). The distance of \( \mathcal{C} \) from the isomorphism class containing a snake and from one containing a star is evidently exactly \( n - k - 1 \). Thus the radius of \( G_n \) is at most \( n - k - 1 = n - \sigma(n) - 1 \).

Conjecture 1. The radius of \( G_n \) 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 \( \mathcal{I}_1 \in \mathcal{I}_n \), \( \mathcal{I}_2 \in \mathcal{I}_n \), \( T_1 \in \mathcal{I}_1 \), \( T_2 \in \mathcal{I}_2 \). Let \( T_1 \), \( T_2 \) be caterpillars and let \( \delta_T(\mathcal{I}_1, \mathcal{I}_2) = k \). 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 \( \delta_T(\mathcal{I}_1, \mathcal{I}_2) = k \), 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 \( n - k \geq 3 \), 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 \( \delta_T(\mathcal{I}_1, \mathcal{I}_2) = k \). 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 \( \mathcal{G}_n \) of \( G_n \) with the property that the distance in \( \mathcal{G}_n \) is the same as in \( G_n \). The diameter of \( \mathcal{G}_n \) is \( n - 3 \).

Now for every positive integer \( k \) we construct a caterpillar \( \bar{T}(k) \). The body of \( \bar{T}(k) \) is a snake of the length \( k - 2 \). The degree of any vertex of this body in \( \bar{T}(k) \) is \( k \). Evidently the number of vertices of \( \bar{T}(k) \) is \( k^2 - 2k + 3 \).
Lemma 2. The caterpillar $T(k)$ 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 $\bar{\rho}$ be the radius of $\mathcal{G}_n$. Then

$$\bar{\rho} \leq n - \bar{\sigma}(n) - 1,$$

where

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

Proof is analogous to that of Theorem 5.

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

In the end we shall compare the distance $\delta_T$ with the distance introduced in [1] on the set of all isomorphism classes of undirected graphs with $n$ vertices. The distance $\delta(G_1, G_2)$ 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 $G_1$ and an induced subgraph belonging to $G_2$.

Theorem 8. For two elements $\mathcal{X}_1, \mathcal{X}_2$ of $\mathcal{G}_n$ for $n \geq 7$ the distances $\delta_T(\mathcal{X}_1, \mathcal{X}_2)$, $\delta(\mathcal{X}_1, \mathcal{X}_2)$ are different in general.

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

$$\delta(\mathcal{G}_1, \mathcal{G}_2) \leq \lceil 3n/2 \rceil - n = \lceil n/2 \rceil < n - 3 = \delta_T(\mathcal{G}_1, \mathcal{G}_2).$$

Reference


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

.130