JOURNAL OF COMBINATORIAL THEORY (B) 17, 48–50 (1974)

A Note on the Metric Properties of Trees*

PETER BUNEMAN*

Communicated by Frank Harary

Received February 21, 1973

By checking the possible configurations of paths which can connect four points x, y, z, t in a tree, it can be seen that the graphical distance [1] must satisfy the inequality:

$$d(x, y) + d(z, t) \leqslant \max \begin{cases} d(x, z) + d(y, t), \\ d(x, t) + d(y, z). \end{cases}$$

We shall refer to this condition as the four-point condition: it is stronger than the triangle inequality (put $z = t$) and is equivalent to saying that of the three sums $d(x, y) + d(z, t)$, $d(x, z) + d(y, t)$, and $d(x, t) + d(y, z)$ two are equal and not less than the third. The four-point condition is also a sufficient condition for a graph to be a tree in the following sense.

THEOREM 1. *A graph is a tree iff it is connected, contains no triangles, and has graphical distance satisfying the four-point condition.*

Proof. Necessity follows immediately, and to prove sufficiency, assume that the graph, T, contains a circuit. Choose a circuit of minimum length p. Since T contains no triangles, $p = 4q + r$ with $q \geqslant 1$ and $0 \leqslant r \leqslant 3$. By the minimality of the circuit, distances between points on it can be measured along paths in the circuit. Therefore, we can choose points x, y, z, t in the circuit such that distances $d(x, y)$, $d(y, z)$, $d(z, t)$, and $d(t, x)$ are all either q or $q + 1$. These points then violate the four-point condition and it follows that there can be no circuit in T.

We can also attach a positive weight λ_e to each edge of a tree and define a new distance d between points x, y of T by

$$d(x, y) = \sum_{e \in E(x, y)} \lambda_e ,$$

where $E(x, y)$ is the set of edges on the path between x and y. We shall call a tree with weighted edges a *weighted* tree and say that it *induces* the

* Present address: Theoretical Psychology Unit, School of Artificial Intelligence, University of Edinburgh, Scotland.

distance d. This new distance is also metric and also satisfies the four-point condition. It also gives us a method for building a tree on any finite set on which a four-point distance is defined. In doing this, we may have to add in new points as nonterminal points of the tree. The construction of these points and the tree are given in the following result.

THEOREM 2. *Let d be a metric on a set S which satisfies the four-point condition. Then there is a weighted tree T which contains the members of S among its points and which induces d.*[1]

Proof. For any metric on a set of order three it is easy to construct the appropriate tree. Our proof is inductive, so we assume that the theorem holds for any set of order less than n.

If S has n members, out of all ordered triples of members of S we select a triple (p, q, r) for which

$$d(p, r) + d(q, r) - d(p, q) \text{ is maximum.}$$

From this, we have for any x in $S - \{p, q\}$,

$$d(x, r) + d(p, q) \leqslant d(x, q) + d(p, r)$$

and

$$d(x, r) + d(p, q) \leqslant d(x, p) + d(q, r).$$

Using the four-point condition on x, p, q, r we get:

$$d(x, q) + d(p, r) = d(x, p) + d(q, r),$$

and similarly for any other member y of $S - \{p, q\}$,

$$d(y, q) + d(p, r) = d(y, p) + d(q, r).$$

Combining these last two gives:

$$d(y, p) + d(x, q) = d(x, p) + d(y, q). \tag{1}$$

We now describe a new object t by giving its dissimilarity from each member of S. The definition of D is extended so that

$$d(t, p) = \tfrac{1}{2}(d(p, q) + d(p, r) - d(q, r)),$$
$$d(t, x) = d(x, p) - d(t, p), \qquad x \neq p.$$

[1] The referee has pointed out that the addition of the constraint that $d(x, y) + d(x, z) - d(y, z)$ should be an even integer gives an alternative formulation of Pereira's [4] conditions for a tree-realizable metric.

It follows that

$$d(p, x) = d(p, t) + d(t, x), \tag{2a}$$

and by using Eq. (1) above that

$$\begin{aligned} d(q, x) &= d(p, x) + d(q, r) - d(p, r) \\ &= d(t, x) + d(t, p) + d(q, r) - d(p, r) \\ &= d(q, t) + d(t, x). \end{aligned} \tag{2b}$$

It also follows from the definition of d on $S \cup \{t\}$ that d still satisfies the four-point condition. We can therefore deduce that d is non-negative; it is also symmetric. Moreover, if $d(t, t') = 0$ for some t in $S - \{p, q\}$, then $d(t, x) = d(t', x)$ for all x in $S - \{p, q\}$ and we can identify t and t'. The removal of $\{p, q\}$ and the addition of t gives us a set with at most $n - 1$ members on which d is a metric satisfying the four-point condition. Now the inductive hypothesis can be used to construct a weighted tree which contains t as a point and induces d. p and q can then be attached by edges of weight $d(p, t)$ and $d(q, t)$, respectively, and Eqs. (2a) and (2b) above ensure that this new weighted tree induces d on S.

One of the problems [2] in automatic classification is to find continuous, well-defined transformations from an arbitrary measure of dissimilarity to an ultrametric. An ultrametric satisfies the condition $d(x, y) \leqslant \max(d(x, z), d(y, z))$, and this is stronger than the four-point condition. Elsewhere [3], a continuous, well-defined transformation has been described which carries a measure of dissimilarity δ to a four-point metric. It involves finding complementary subsets S^1, S^2 of the set on which δ is defined, which satisfy:

$$\delta(x, z) + \delta(y, t) - \delta(x, y) - \delta(z, t) \geqslant 0$$

for all $x, y \in S^1$, and $z, t \in S^2$. A drawback to this method is that it is sensitive to "large" values of δ and can produce uninteresting trees from a dissimilarity which deviates even a little from the four-point condition. It would be interesting to know if there are other such transformations.

References

1. F. Harary, "Graph Theory," Addison Wesley, 1969.
2. C. J. Jardine, N. Jardine, and R. Sibson, *Math. Biosci.* **1** (1967), 173–179.
3. O. P. Buneman, "Mathematics in the Archaeological and Historical Sciences" (F. R. Hodson, D. G. Kendall, and P. Tautu, eds.), pp. 387–395, Edinburgh University Press, 1971.
4. J. M. S. Pereira, *J. Combinatorial Theory* **6** (1969), 303–310.