The path-variance problem on tree networks

Teresa Cáceres\textsuperscript{a}, M. Cruz López-de-los-Mozos\textsuperscript{a}, Juan A. Mesa\textsuperscript{b}

\textsuperscript{a}Departamento de Matemática Aplicada I, E.T.S. de Ingeniería Informática, Avda. Reina Mercedes, s/n, Sevilla 41012, Spain
\textsuperscript{b}Departamento de Matemática Aplicada II, Universidad de Sevilla, E.T.S. de Ingeniería Informática, Avda. Reina Mercedes, s/n, Sevilla 41012, Spain

Received 15 December 2000; received in revised form 25 June 2002; accepted 19 September 2003

Abstract

Extensive facility location models on graphs deal with the location of a special type of subgraphs such as paths, trees or cycles and can be considered as extensions of classical point location models. Variance is one of the measures applied in models in which some equality requirement is imposed. In this paper the problem of locating a minimum variance path in a tree network is addressed, and an $O(n^2 \log n)$ time algorithm is proposed.

Keywords: Location on networks; Equality; Extensive facilities

1. Introduction

During the last two decades there has been increasing attention paid to the application of balancing or equality objectives in location analysis. Papers on equality location models can be roughly classified into two groups: First, those dealing with general aspects such as how to measure equality, how to define and what properties should have equality measures ([2,3,10]), and secondly, those oriented to design efficient algorithms for solving the corresponding problems ([5,8,9]).

On the other hand, when the facility to be located is too large to be modelled as a point, extensive facility location models arise. Several papers have investigated the problem of locating, on a network, a path, tree or other types of facilities which may not be considered as points (see [11,7] for a survey of this literature). Examples of problems in which structures instead of points are required include the location of pipelines, evacuation routes, mass transit routes or routing a highway through a road network. For the optimal selection of a site in which to locate an extensive facility the criteria used are, in almost all cases, the minimax criterion, the minisum criterion or a covering model, (see [6,12–16]). In addition, although most of these problems can be solved in polynomial time in tree networks, they are NP-hard in cyclic networks (see [4] for an analysis of the complexity of these problems). For this reason the main effort has been orientated to design efficient algorithms to solve the aforementioned problems in tree networks. In particular the path-center and the path-median (or core) of a tree (which arise from the application of the minimax criterion and the minisum criterion, respectively) can be found by means of linear time algorithms (see [15,13]).

The above discussion shows that the equity (as suitable objective) has not been considered in the criteria used hitherto in the location of extensive facilities. However, in certain public sector problems, an equity criterion is needed to generate acceptable decisions. The variance of distance travelled by all customers to the facility has been an equity criteria widely studied in point location problems. The purpose of this work is to apply this criterion for locating a path-shaped facility in a tree network. This problem, which will be called the Path-Variance Problem, will be studied in the continuous version in which partial edges are allowed, that is, when the extreme points of the path can be interior points of an edge. In
general these extreme points will belong to different edges. However, the case in which both points are in the same edge will be analyzed separately, since this gives rise to the particular case in which the path degenerates to a single point.

The remainder of the paper is organized as follows. In the next section the formulation of the problem is introduced and properties of the objective function when the extreme points belong to different edges are studied. Section 3 is devoted to the case in which the extreme points of the path belong to the same edge. In order to design the algorithm for the problem, a data structure to describe the tree is incorporated at the beginning of Section 4. In that section the algorithm is also proposed, and its complexity is discussed.

2. The path-variance problem on different edges

Let \( T = (V, E) \) be an undirected tree network with vertex set \( V = \{v_1, \ldots, v_n\} \), edge set \( E \), and edge lengths \( \{l_j : e_j \in E\} \). Denote by \( \mathcal{F} \) the set of points in \( T \) which comprises both set \( V \) and the set of all the points on edges. We will consider paths not only containing complete edges but also partial edges, and they will be denoted as \( P(x_1, x_2) \) in which \( x_1, x_2 \in \mathcal{F} \) are the end points of the path. If \( P(x_1, x_2) \subseteq \mathcal{F} \) is a path, then \( d(x, P) = \min_{y \in P} d(x, y) \), where \( d(x, y) : x, y \in \mathcal{F} \), is the distance between points given by the length of the only path joining \( x \) and \( y \), which makes \( \mathcal{F} \) a metric space.

We suppose that demand originates at the vertices of the tree, and we associate a positive weight \( w_i \) with each vertex \( v_i \in V \). Without loss of generality we may assume that \( \sum_{i=1}^{n} w_i = 1 \) and interpret each \( w_i \) as the fraction of demand originated at \( v_i \). For a subset of vertices \( V' \subseteq V \) let \( W(V') = \sum_{v \in V'} w_i \) be the weight of \( V' \), and let \( W(P) = \sum_{v \in V \setminus V'} w_i \) be the weight of the complement of \( V' \) in \( V \).

For any point \( x \in \mathcal{F} \), the median function (distanceum) \( z_{\mathcal{F}}(x) \) of \( \mathcal{F} \) over \( V' \subseteq V \) is given by \( z_{\mathcal{F}}(x, V') = \min_{v \in V'} w_i d(v, x) \). Similarly the median function of a path \( P(x_1, x_2) \) in \( \mathcal{F} \) is given by \( z_{\mathcal{F}}(P(x_1, x_2), V') = \sum_{v \in V'} w_i d(v, P(x_1, x_2)) \). For simplicity, when \( V' = V \) we will use \( z_{\mathcal{F}}(x, V) \) and \( z_{\mathcal{F}}(P(x_1, x_2), V) \) instead of \( z_{\mathcal{F}}(x, V') \) and \( z_{\mathcal{F}}(P(x_1, x_2), V') \), respectively.

The variance function for a path \( P(x_1, x_2) \) is defined as

\[
z_m(P(x_1, x_2)) = \sum_{i=1}^{n} w_i [d(v_i, P(x_1, x_2)) - z_m(P(x_1, x_2))]^2.
\]

The Continuous Path-Variance Problem (CPVP) consists of finding points \( x_1, x_2 \) such that the path joining these points minimises the function \( z_m \) on all the paths \( P \in \mathcal{F} \).

Given a pair of edges \( e_j, e_k \) the Continuous Path-Variance Problem Restricted to \( (e_j, e_k) \), (CPVP)_\{\j\k\}, is that of minimising the restricted function \( z_{\mathcal{F}}(P(x_1, x_2)) \), where \( x_1 \in e_j, x_2 \in e_k \), whose domain is \([0, l_j] \times [0, l_k]\) respectively. Consequently a solution of the problem (CPVP)_\{\j\k\} is a pair \((x_j^*, x_k^*) \in [0, l_j] \times [0, l_k]\) such that

\[
z_m(P(x_j^*, x_k^*)) \leq z_m(P(x_j, x_k)), \quad \forall (x_j, x_k) \in [0, l_j] \times [0, l_k].
\]

As usual when an edge \( e_j \) is fixed, any real number \( x \in [0, l_j] \) will denote the point on the edge whose exact location is determined by its distance (along the edge) from a prescribed endpoint of the edge. In order to avoid a decomposition into two cases, and therefore the subsequent case analysis, the reference vertices to measure distances between any two points \( x_1 \in e_j = (u_j, v_j) \) and \( x_2 \in e_k = (u_k, v_k) \) will be the farthest vertices among those of both edges. We will suppose that \( u_j \) and \( u_k \) are these farthest vertices, consequently \( x_1 \in e_j \) means the length of subedge \((u_j, x_1)\) and \( x_2 \in e_k \) means the length of subedge \((v_k, x_2)\).

For a vertex \( u \) of edges \( e_j, e_k \), let \( V_u \) denote the vertex set of the subtree \( T_u \) rooted at \( u \) which does not contain the edge \( e_j \) or \( e_k \) which is incident to \( u \). \( W(V_u) = \sum_{v \in V_u} w_i \) is the total weight of \( V_u \) and \( W(T_u) = 1 - W(V_u) \) is the weight of the complementary vertex set. In accordance with this notation, \( T_{u_j} \) and \( T_{u_k} \) will represent the subtrees rooted at the farthest vertices \( u_j \) and \( v_k \) respectively, and do not contain the edges \( e_j \) and \( e_k \) (and therefore neither do they contain the path \( P(x_1, x_2) \)).

Likewise we denote by \( V(P(x_1, x_2)) \) the vertex set of the path \( P(x_1, x_2) \) and by \( T_1, T_2, \ldots, T_r \) the connected components (subtrees) that result from the deletion of \( V(P(x_1, x_2)) \) and the corresponding incident edges from the tree \( T \setminus (T_{u_j} \cup T_{u_k}) \).

Then the median function for the path \( P(x_1, x_2) \) is

\[
z_m(P(x_1, x_2)) = \sum_{v_i \in P} w_i d(v_i, P(x_1, x_2)) = \sum_{v_i \in V_{u_j}} w_i [d(v_i, u_j) + x_1] + \sum_{v_i \in V_{u_k}} w_i [d(v_i, v_k) + x_2]
\]

\[
= z_m(P(u_j, v_j)) + W(V_{u_j}) x_1 + W(V_{u_k}) x_2.
\]
In the following we will denote by \( z_m(P) \) the value \( z_m(P(u_j, v_k)) \).

The variance of the path \( P(x_1, x_2) \) can be decomposed in the following manner:

\[
z_z(P(x_1, x_2)) = \sum_{v_i \in V_k} w_i [(d(v_i, u_j) - z_m(P))^2 + (W(V_{u_j}))^2 x_1^2]
\]

\[
+ (W(V_{u_j}))^2 x_2^2 + 2(d(v_i, u_j) - z_m(P)) W(V_{u_j}) x_1
\]

\[
- 2(d(v_i, u_j) - z_m(P)) W(V_{u_j}) x_2 - 2W(V_{u_j}) W(V_{x_2}) x_1 x_2 \]

\[
+ \sum_{k=1}^{r} \left( \sum_{v_i \in V_k} w_i [(d(v_i, P) - z_m(P))^2 + (W(V_{P}))^2 x_1^2 + (W(V_{P}))^2 x_2^2]
\]

\[
- 2(d(v_i, P) - z_m(P)) W(V_{P}) x_1 - 2(d(v_i, P) - z_m(P)) W(V_{P}) x_2
\]

\[
+ 2W(V_{P}) W(V_{x_2}) x_1 x_2 \right)
\]

\[
+ \sum_{v_i \in V_k} w_i [(d(v_i, v_k) - z_m(P))^2 + (W(V_{P}))^2 x_1^2 + (W(V_{P}))^2 x_2^2]
\]

\[
+ 2(d(v_i, v_k) - z_m(P)) W(V_{P}) x_2 - 2(d(v_i, v_k) - z_m(P)) W(V_{P}) x_1
\]

\[
- 2W(V_{P}) W(V_{x_2}) x_1 x_2 \].
\]

By associating terms with equal degree,

\[
z_z(P(x_1, x_2)) = W(V_{u_j}) W(V_{u_j}) x_1^2 + W(V_{x_2}) W(V_{x_2}) x_2^2
\]

\[
- 2W(V_{u_j}) W(V_{x_2}) x_1 x_2
\]

\[
+ 2(z_m(u_j, v_i) - W(V_{u_j}) z_m(P)) x_1
\]

\[
+ 2(z_m(v_k, v_i) - W(V_{x_2}) z_m(P)) x_2 + z_z(P).
\]

By imposing the necessary conditions for stationary points the following system results:

\[
W(V_{u_j}) W(V_{u_j}) x_1 - W(V_{u_j}) W(V_{x_2}) x_2 + (z_m(u_j, v_i) - W(V_{u_j}) z_m(P)) = 0,
\]

\[
W(V_{x_2}) W(V_{u_j}) x_2 - W(V_{x_2}) W(V_{x_2}) x_1 + (z_m(v_k, v_i) - W(V_{x_2}) z_m(P)) = 0.
\]

Let \( W(P) = \sum_{v_i \in V(P(x_1, x_2))} w_i \) be the weight of the vertex set of the path \( P(x_1, x_2) \). Clearly \( W(P) > 0 \) since the set \( V(P(x_1, x_2)) \) is not empty when the extreme points \( x_1, x_2 \) of the path belong to different edges. By simplifying the unique solution of the last system, we obtain

\[
x_1 = \frac{1}{W(P)} \left( z_m(P) - z_m(v_k, v_i) - z_m(u_j, v_i) W(V_{x_2}) \left( \frac{W(\overline{V}_{u_j})}{W(V_{u_j})} \right) \right)
\]

\[
x_2 = \frac{1}{W(P)} \left( z_m(P) - z_m(u_j, v_i) - z_m(v_k, v_i) W(V_{u_j}) \left( \frac{W(\overline{V}_{x_2})}{W(V_{x_2})} \right) \right).
\]

Furthermore, \( 0 \leq x_1 \leq l_j \) and \( 0 \leq x_2 \leq l_k \) must hold.

The Hessian matrix is

\[
H = \begin{pmatrix}
2W(V_{u_j}) W(V_{u_j}) & -W(V_{u_j}) W(V_{x_2}) \\
-W(V_{x_2}) W(V_{x_2}) & 2W(V_{x_2}) W(V_{x_2})
\end{pmatrix}.
\]
Since \( W(\mathcal{T}_{u_j}) = W(P) + W(V_{u_j}) \) and \( W(\mathcal{T}_{u_k}) = W(P) + W(V_{u_k}) \), then
\[
|H| = W(V_{u_j})W(V_{v_j})(4W(\mathcal{T}_{u_j})W(\mathcal{T}_{v_j}) - W(V_{u_j})W(V_{v_j})) > 0
\]
and also \( 2W(V_{u_j})W(\mathcal{T}_{u_j}) > 0 \).

Therefore, the Hessian matrix is definite positive, which guarantees that the variance function is strictly convex on the compact set \([0, l_j] \times [0, l_k] \) and the stationary point given by (2) is the minimum, which implies that it is the solution \((x_1^*, x_2^*)\) of the restricted problem \((CPVP)_{u_j}\).

When the minimum \((x_1, x_2)\) does not belong to the above compact set then the optimum is reached in the boundary. There are several cases.

1. If \( x_1 \leq 0 \) and \( 0 \leq x_2 \leq l_k \) then the minimum \((x_1^*, x_2^*)\) of the restricted problem is reached at the point
\[
\left(0, \frac{W(V_{u_j})z_m(P) - z_m(v_k, V_{v_j})}{W(\mathcal{T}_{v_j})W(V_{v_j})}\right).
\]
2. If \( x_1 \geq l_j \) and \( 0 \leq x_2 \leq l_k \) then the optimum is
\[
\left(l_j, \frac{W(V_{u_j})z_m(P) + W(V_{u_j})l_j - z_m(v_k, V_{v_j})}{W(V_{v_j})W(\mathcal{T}_{v_j})}\right).
\]
3. If \( x_2 \leq 0 \) and \( 0 \leq x_1 \leq l_j \) then the minimum is given by
\[
\left(\frac{W(V_{u_j})z_m(P) - z_m(u_j, V_{u_j})}{W(V_{u_j})W(\mathcal{T}_{u_j})}, 0\right).
\]
4. If \( x_2 \geq l_k \) and \( 0 \leq x_1 \leq l_j \) then
\[
\left(\frac{W(V_{u_j})z_m(P) - z_m(u_j, V_{u_j})}{W(V_{u_j})W(\mathcal{T}_{u_j})}, l_k\right).
\]
5. Finally,
\[
(x_1^*, x_2^*) = \begin{cases} 
(l_j, 0) & \text{if } x_1 \geq l_j \text{ and } x_2 \leq 0, \\
(0, 0) & \text{if } x_1 \leq 0 \text{ and } x_2 \leq 0, \\
(l_j, l_k) & \text{if } x_1 \geq l_j \text{ and } x_2 \geq l_k, \\
(0, l_k) & \text{if } x_1 \leq 0 \text{ and } x_2 \geq l_k.
\end{cases}
\]

3. The path-variance problem on an edge

In this section the path location problem, in which the two ends of the path are in the same edge, is considered. We are looking for the path that minimises the variance function \( z_s(P(x_1, x_2)) \) among those paths \( P(x_1, x_2) \) such that \( x_1, x_2 \in e_j = (u_j, v_j) \). Following the notation introduced in the previous section, \( u_j \) and \( v_j \) are now the farthest vertices. Consequently \( x_1, x_2 \) represent the lengths of the subedges \((u_j, x_1)\) and \((v_j, x_2)\), respectively, and if \( V_{u_j}, V_{v_j} \) are the vertex sets of the subtrees \( T_{u_j}, T_{v_j} \) rooted at \( u_j, v_j \), respectively, which do not contain the edge \( e_j \), then they are complementary vertex sets, i.e. \( V_{u_j} \cup V_{v_j} = V \). This implies that \( W(P) = 0 \), which justifies a separate study of this case.

The expression for the median function is the same as in the previous case, \( z_m(P(x_1, x_2)) = z_m(P(u_j, v_j)) + W(V_{u_j})x_1 + W(V_{v_j})x_2 \). By using the simplified notation \( z_m(P) \) and \( z_s(P) \) instead of \( z_m(P(u_j, v_j)) \) and \( z_s(P(u_j, v_j)) \), respectively, and taking into account that \( \mathcal{T}_{u_j} = V_{u_j} \), the variance function can be written as follows:
\[
z_s(P(x_1, x_2)) = W(V_{u_j})W(V_{v_j})x_1^2 + W(V_{u_j})W(V_{v_j})x_2^2
\]
\[
- 2W(V_{u_j})W(V_{v_j})x_1x_2 + 2\{z_m(u_j, V_{u_j}) - W(V_{u_j})z_m(P)\}x_1
\]
\[
+ 2\{z_m(v_j, V_{v_j}) - W(V_{v_j})z_m(P)\}x_2 + z_s(P) = W(V_{u_j})W(V_{v_j})[x_1 - x_2]^2
\]
\[
+ 2\{z_m(u_j, V_{u_j})W(V_{u_j}) - z_m(v_j, V_{v_j})W(V_{v_j})\}[x_1 - x_2] + z_s(P).
\]
It can be seen that when \( x_1 = x_2 \) (i.e. the path degenerates to the medium point of the edge), then the value \( z_s(P(x_1, x_2)) \) coincides with the variance \( z_s(P) \) of the complete edge. Moreover, the edge variance also coincides with the variance of any path centered at the medium point and contained in the edge. The necessary conditions for stationary points give rise to the following system:

\[
W(V_{u_j})W(V_{v_j})(x_1 - x_2) + (z_m(u_j, V_{u_j}) - W(V_{u_j})z_m(P)) = 0,
\]

\[
W(V_{v_j})W(V_{u_j})(x_2 - x_1) + (z_m(v_j, V_{v_j}) - W(V_{v_j})z_m(P)) = 0,
\]

in which \( z_m(u_j, V_{u_j}) - W(V_{u_j})z_m(P) = W(V_{v_j})z_m(P) - z_m(v_j, V_{v_j}) \), since \( z_m(P) = z_m(u_j, V_{u_j}) + z_m(v_j, V_{v_j}) \) and \( W(V_{u_j}) + W(V_{v_j}) = 1 \). This implies that the system is undetermined compatible, and it provides the only condition given by

\[
x_1 - x_2 = \frac{z_m(P)}{W(V_{v_j})} - \frac{z_m(u_j, V_{u_j})}{W(V_{u_j})} - \frac{z_m(v_j, V_{v_j})}{W(V_{v_j})} = K,
\]

or equivalently

\[
x_1 - x_2 = \frac{z_m(v_j, V_{v_j})}{W(V_{v_j})} - \frac{z_m(u_j, V_{u_j})}{W(V_{u_j})} = K,
\]

with \( 0 \leq x_1 \leq l_j, \ 0 \leq x_2 \leq l_j, \) and \( 0 \leq x_1 + x_2 \leq l_j \).

Since the Hessian matrix is also definite positive the stationary points are minima. In accordance with these expressions, the following cases can be considered (see Fig. 1).

**Case 1:** If \(-l_j < K < l_j\) then any path for which

\[
x_1 - x_2 = K, \quad 0 < x_1 + x_2 < l_j, \quad 0 < x_1 \leq x_2 \leq l_j
\]

is a path of minimum variance whose variance value is:

\[
z_s(P(x_1, x_2)) = z_s(P) + K^2W(V_{u_j})W(V_{v_j})
\]

\[
+ 2K\left\{W(V_{v_j})z_m(u_j, V_{u_j}) - W(V_{u_j})z_m(v_j, V_{v_j})\right\}
\]

\[
= z_s(P) + \frac{[z_m(v_j, V_{v_j})W(V_{u_j}) - z_m(u_j, V_{u_j})W(V_{v_j})]^2}{W(V_{u_j})W(V_{v_j})}
\]

\[
= z_s(P) - K^2W(V_{u_j})W(V_{v_j}).
\]

Let us note that the point \((x_1, x_2) \in [0, l_j] \times [0, l_j]\) such that

\[
x_1 + x_2 = l_j, \quad x_1 - x_2 = K
\]

is the point of minimum variance of this edge.

**Case 2:** If \( K \geq l_j \) then the path degenerates to the point \( v_j \), which corresponds to the pair \((l_j, 0)\), and also is the point of minimum variance of the edge.

**Case 3:** If \( K \leq -l_j \) then the path reduces to the point of minimum variance of the edge, \( u_j \).
In this case, the vertices belonging to the edge \( e_k \) traversal of \( T \) can be obtained from that used in Section 2, by applying the simple change \( P(j) \) (if \( j \neq k \)) and \( P(k) \) (if \( j = k \)), requires the previous computation of a set of auxiliary values associated to each vertex \( u \) as well as the evaluation of the function \( z_m \) on the path joining the farthest vertices of the edges \( e_j \) and \( e_k \). In order to obtain all these values the recursive procedure designed by Maimon [9] will be applied.

Let \( v_r \) be the root with sons \( v_1, \ldots, v_s \) (where as we have already seen, \( v_r \) can be either \( v_{i(k)} \) or \( u_k \)), (i.e. there exists an edge \((v_i, v_r)\) \( \in E \) for each \( i = 1, \ldots, s \)). A postorder traversal of \( T \) is defined recursively as follows:

1. Visit in postorder the subtrees with roots \( v_1, \ldots, v_s \) in that order.
2. Visit the root \( v_r \).

The postorder defined in \( T \) induces the following order of edges of \( E \):

\[
(u_1, v_{(i1)}), (u_2, v_{(i2)}), \ldots, (u_{s-1}, v_r),
\]

where each \( v_{(i)} \) is the father of \( u_i \) in the postorder, and for each two edges \( e_j = (u_j, v_{(i)}) \) and \( e_k = (u_k, v_{(k)}) \), \( k > j \) means that \( e_k \) can be (not necessarily) a common ancestor of the remaining three vertices.

This fact provides an identification of the farthest vertices of pair \( e_j = (u_j, v_{(i)}) \), \( e_k = (u_k, v_{(k)}) \) (with \( k > j \)) as follows (see Fig. 2).

**Case (i):** If \( v_{(i)} \) is a descendant of \( u_k \) (that is, \( v_{(i)} \in V_{uk} \) in the defined postorder), then \( u_j, v_{(i)} \) are the farthest vertices. In this case, the vertices belonging to \( P = P(u, v_{(i)}) \) can be found by successively adding the vertex father of each vertex son (starting with the first son \( u_j \), and ending with the last father \( v_{(i)} \)).

**Case (ii):** Otherwise, the farthest vertices are \( u_j, u_k \). In this case, \( u_j, u_k \) have a common ancestor \( v^* \), which can be either one of the fathers \( v_{(i)}, v_{(k)} \) (or both, if they coincide), or a vertex not in \( e_j, e_k \). In this last case, the path \( P \) joining \( u_j, u_k \) can be found by a binary search among the fathers of the vertices \( u_j, u_k \) as follows: starting with \( u_j \) and \( u_k \), the successive fathers are tested (and added to path \( P = P(u_j, u_k) \) until a common father \( v^* \) is found).

If \( v_k \) denotes the farthest vertex from \( u_k \) (where as we have already seen, \( v_k \) can be either \( v_{i(k)} \) or \( u_k \)), the point \( x_2 \) in the edge \( e_k \) is identified by its distance from \( v_k \). Note that in case (i) the identification of the second point when using the postorder can be obtained from that used in Section 2, by applying the simple change \( x'_2 = l_k - x_2 \). However, the point \( x_1 \) in the edge \( e_j \) is always identified by its distance from the lower vertex \( u_j \) of the edge.

If \( T_u = (V_u, E_u) \) denotes (in the postorder) the subtree rooted in each \( u \in V \), the necessary vertex information for computing the local optimum \((x_1^*, x_2^*)\) can be determined in a preprocessing phase, in which the tree is twice traversed following the postorder (toward the root of the tree and conversely). By applying the recursive relationships of Maimon [9], at the end of this phase the aforementioned vertex information is available. This information (accumulated in a vector of auxiliary values associated to each vertex) is:

\[
W(V_u), z_m(u, V_u), z_m(u), z_m^2(u, V_u), z_m^2(u),
\]

where \( z_m^2(u, V_u) = \sum_{v \in V_u} w_d(v, u)^2 \), and \( z_m^2(u) = z_m^2(u, V) \). It is easy to see that for a path \( P \)

\[
z_d(P) = \sum_{v \in P} w_d(v, P)^2 - \left( \sum_{v \in P} w_d(v, P) \right)^2 = z_m^2(P) - (z_m(P))^2.
\]

Therefore the aforementioned auxiliary values will be necessary for computing \( z_d(P) \).
If \( \overline{V}_u = V \setminus V_u \) and \( S_u \) is the set of sons of \( u \) and \( v \) its father, the set \( \mathcal{C}(u) = S_u \cup \{v\} \) contains all the nearest vertices (from \( u \)) in the connected components obtained by deleting \( u \) and the incident edges. The preprocessing phase also provides the values associated to the connected components whose nearest vertices belong to \( S_u \):

\[
\{ W(V_u), z_m(u), V_u \}, z_m^{(2)}(u), V_u \quad \forall u \in S_u
\]
as well as the values associated to the connected component father:

\[
W(\overline{V}_u) = 1 - W(V_u), \\
z_m(v, \overline{V}_u) = z_m(u) - z_m(u, V_u) - l_m(1 - W(V_u)), \\
z_m^{(2)}(v, \overline{V}_u) = z_m^{(2)}(u) - z_m^{(2)}(u, V_u) - l^2_m W(\overline{V}_u) - 2l_m z_m(v, \overline{V}_u).
\]

The knowledge of these auxiliary values allows us to obtain \( z_m(P) \) and \( z_m^{(2)}(P) \) by means of a progressive procedure. Let \( P = P(u_j, v_k) \) be the path joining the farthest vertices in the pair of edges \( \{e_j, e_k\} \), \( k > j \), let \( V(P) \) be the vertex set of the path \( P \) and let \( \mathcal{C}(u_j) = \{u_j, \ldots, u_k\}, \mathcal{C}(v_k) = \{v_k, \ldots, v_j\} \) be the respective nearest vertices in the connected components associated to \( u_j, v_k \). Initially

\[
z_m(P) = z_m(u_j) - \sum_{u_i \in \mathcal{C}(u_j) \cap V(P)} z_m(u_i, T_{u_i}) + z_m(v_k) - \sum_{v_i \in \mathcal{C}(v_k) \cap V(P)} z_m(v_i, T_{v_i}).
\]

When a vertex \( v' \) is added to \( P \) according to the aforementioned described cases (i) or (ii), the values of \( z_m(P) \) and \( z_m^{(2)}(P) \) are successively updated by adding the corresponding auxiliary values of \( v' \) in the connected components whose nearest vertices are not in \( P \):

\[
z_m(P) \leftarrow z_m(P) + z_m(v') - \sum_{v_i' \in \mathcal{C}(v') \cap V(P)} z_m(v_i', T_{v_i'}). 
\]

Note that the cardinality of the set \( \mathcal{C}(v') \cap V(P) \) is bounded by 2 (since there exists at most two vertices in \( P \) which are adjacent to \( v' \)). Likewise, \( z_m^{(2)}(P) \) is obtained by the same procedure by considering \( z_m^{(2)}(v') \) instead of \( z_m \).

Finally, before describing the pseudocode of the algorithm, it is necessary to make some observations with respect to the value \( W(\overline{V}_v) \) which appears in the expression (2). For a given pair of edges \( e_j = (u_j, v_{(j)}) \), \( e_k = (u_k, v_{(k)}) \) whose farthest vertices are \( u_j, v_k \), for the case (ii) \( W(\overline{V}_{v_k}) \) coincides with the auxiliary value provided by the postorder, that is: \( W(\overline{V}_v) = 1 - W(V_v) \). However, in the case (i) such a value represents the weight of the connected component associated to \( v_k \) whose nearest vertex is \( u_k \), that is, \( u_k \in \mathcal{C}(v_k) \cap V(P) \). Therefore, in the case (i), \( W(\overline{V}_v) = W(V(T_{u_k})) \), and \( W(V_{v_k}) = 1 - W(\overline{V}_{v_k}) \). Thus, the pseudocode of the algorithm is as follows:

**Input:** A tree network given by a postorder traversal.

**Preprocessing phase**

**For** \( (u \in V) \) **do**

Compute the auxiliary values.

**endfor**

Let \( Z \) be a suitable large number.

**Main step**

**For** \( (j = 1 \text{ to } n - 1) \) **do**

**For** \( (k = j \text{ to } n - 1) \) **do**

If \( (k > j) \) **then**

By a binary search determine the vertex set of 
\( P = P(u_j, v_k) \) and simultaneously compute 
\( z_m(P) \) and \( z_m^{(2)}(P) \).

Compute \( z_m(P) = z_m^{(2)}(P) - (z_m(P))^2 \).

Compute the point \((x_{1,1}, x_{1,2})\) (by applying (2) and subsequent relationships).

**Otherwise** \( (k = j) \) **then**

Compute \((x_{1,1}^*, x_{1,2}^*)\) (by applying (3) and subsequent relationships).

**endif**
Compute $z_s = z_s(P(x^*_1, x^*_2))$.

If $(z_s < Z)$ then

\[ P^* \leftarrow P(x^*_1, x^*_2), \quad Z \leftarrow z_s. \]

endif

endfor

Output: The path $P^* = P(x^*_1, x^*_2)$ that minimises the variance and $Z = z_s(P^*)$.

For each considered path, obtaining the optimum requires identifying and adding a set of auxiliary values associated to the corresponding nodes of the path. By using the postorder structure, such nodes (and their auxiliary values) are added by means of a binary search process over the set of fathers. Since the complexity of the binary search is $O(\log n)$ (see [1]) and $n^2$ paths are tested, the overall complexity of the algorithm is $O(n^2 \log n)$ time.

Acknowledgements

The authors are partially supported by Project: MCyT BFM2000-1052-C02-01.

References