Discrete Mathematics, Algorithms and Applications
Vol. 5, No. 3 (2013) 1350010 (9 pages)

DOI: 10.1142/S1793830913500109

DISTANCE PRESERVING SUBTREES IN MINIMUM AVERAGE DISTANCE SPANNING TREES

LAURENT LYAUDET[*,§], PAULIN MELATAGIA YONTA[†,‡,¶], MAURICE TCHUENTE[*,†,‡,‖] and RENÉ NDOUNDAM[†,**]

[*]*Ecole Normale Supérieure de Lyon, Institut des Systèmes Complexes 46 Allée d'Italie, 69364 Lyon cedex 07, France*

[†]*Université de Yaoundé I, UMI 209, UMMISCO, Faculté des Sciences Département d'Informatique, B.P. 812 Yaoundé, Cameroun*

[‡]*LIRIMA, Equipe IDASCO Université de Yaoundé 1, B.P. 337 Yaoundé, Cameroun*

[§]*laurent.lyaudet@gmail.com*
[¶]*paulinyonta@gmail.com*
[‖]*maurice.tchuente@ens-lyon.fr*
[**]*ndoundam@yahoo.com*

Received 25 July 2012
Revised 23 November 2012
Accepted 15 December 2012
Published 6 August 2013

Given an undirected graph $G = (V, E)$ with n vertices and a positive length $w(e)$ on each edge $e \in E$, we consider Minimum Average Distance (MAD) spanning trees i.e., trees that minimize the path length summed over all pairs of vertices. One of the first results on this problem is due to Wong who showed in 1980 that a Distance Preserving (DP) spanning tree rooted at the median of G is a 2-approximate solution. On the other hand, Dankelmann has exhibited in 2000 a class of graphs where no MAD spanning tree is distance preserving from a vertex. We establish here a new relation between MAD and DP trees in the particular case where the lengths are integers. We show that in a MAD spanning tree of G, each subtree $H' = (V', E')$ consisting of a vertex $\overline{r}$ and the union of branches of $\overline{r}$ that are each of size less than or equal to $\sqrt{\frac{n}{2w^+}}$, where w^+ is the maximum edge-length in G, is a distance preserving spanning tree of the subgraph of G induced by V'.

Keywords: Undirected graph; minimum average distance spanning tree; distance preserving.

Mathematics Subject Classification: 05C12, 05C05, 05C75

1. Introduction

Consider an undirected and connected graph $G = (V, E)$, where $V = \{1, 2, \ldots, n\}$ is the set of vertices and E is the set of edges. Hereafter $V(G)$ and $E(G)$ will also be used to denote respectively the set of vertices and the set of edges of a graph G.

[¶]Corresponding author.

A function w assigns to each edge $e = \{x, y\}$ a positive length denoted $w(e)$ or $w(x, y)$. The length of a path is the sum of the lengths of the edges along this path. For any pair of vertices $\{x, y\} \subset V^2$, $d_G(x, y)$ denotes the distance from x to y, i.e., the length of the shortest path from x to y in G. Hereafter, we denote by $D_G(X, Y)$ the sum $\sum_{x\in X}\sum_{y\in Y} d_G(x, y)$, for $X, Y \subseteq V$ and $D(G) = D_G(V, V)$. A Minimum Average Distance (MAD) spanning tree of G is a spanning tree T that minimizes $\frac{D(T)}{n(n-1)}$. Clearly, such a tree also minimizes $D(T)$. The notation $G[X]$ will be used in this paper to represent the subgraph of G induced by $X \subseteq V$.

Definition 1.1. Let T be a tree and let $x, y \in V(T)$. A branch of vertex x in the tree T is a connected component of the forest $T\backslash x$ obtained from T by deleting x. The root of a branch of vertex x is the vertex of this branch which is a neighbour of x in T. The subtree of a vertex y with respect to x is denoted $T_x(y)$ and is obtained from T by removing the branch of y that contains x.

Definition 1.2. A spanning tree T of a graph $G = (V, E)$ is said to be Distance Preserving (DP) if there exists a vertex r such that $d_T(r, x) = d_G(r, x)$ for all $x \in V$. T is said to be distance preserving from r.

Wong has shown in [4] that a DP spanning tree rooted at the median of G is a 2-approximate solution of the MAD spanning tree problem. On the other hand, Dankelmann has exhibited in [1] a class of graphs where no MAD spanning tree is distance preserving. We establish here a new relation between MAD and DP trees in the particular case where the lengths are integers. We show that in a MAD spanning tree of G, each subtree $H' = (V', E')$ consisting of the union of branches of a vertex $\overline{r}$ that are each of size less than or equal to $\sqrt{\frac{n}{2w^+}}$, where w^+ is the maximum edge-length in G, is distance preserving with respect to the subgraph of G induced by V'. This result establishes a relation between MAD spanning trees and DP subtrees which have been extensively used to build approximation algorithms for the MAD spanning tree problem [4–6].

The remainder of the paper is organized as follows. In Sec. 2, the results of Wong and Dankelmann concerning the relations between MAD and DP spanning trees, together with the proofs are presented. In Sec. 3, we give a simple result which says that if G is an undirected graph with unit edge-lengths, then in MAD trees, any branch of size less than or equal to $\sqrt[3]{n}$ is distance preserving from its root. This result is improved in Sec. 4, where we present the main result of this paper. Section 5 is devoted to concluding remarks.

2. Previous Results

In this section, we recall two results concerning the relation between MAD and DP spanning trees.

Definition 2.1. A median of G is a vertex m that minimizes $\sum_{x\in V} d_G(m, x)$.

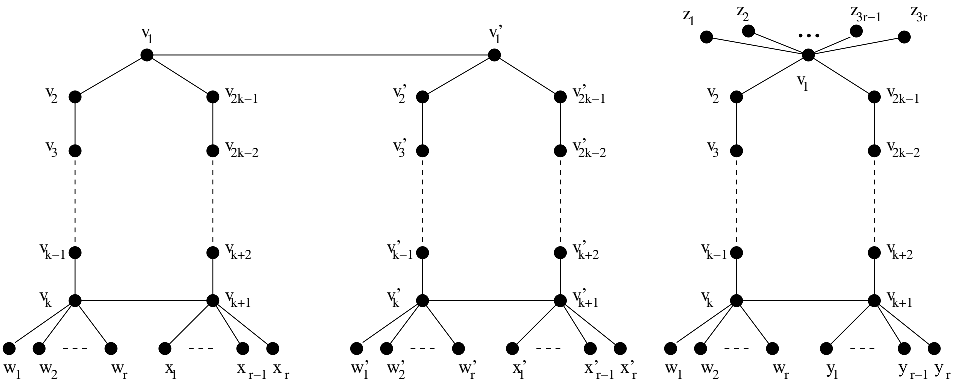

Fig. 1. Counter examples for Entringer's questions.

Proposition 2.2. (*Wong* [4]) *A DP spanning tree $\hat{T}$ from a median m is a 2-approximation of the MAD spanning tree $\overline{T}$ of G, i.e., $D(\hat{T}) \leq 2D(\overline{T})$.*

Proposition 2.2 establishes a positive relationship between MAD and DP spanning trees. Following this result, Entringer [2] has asked the following questions:

(1) Does every graph has a MAD spanning tree that is distance preserving?
(2) If a graph G has a MAD spanning tree $\hat{T}$ that is distance preserving, is it distance preserving from a median?

Consider the graph G of Fig. 1(a) having edges of length 1. Assume that a MAD spanning tree $\overline{T} = (V, \overline{E})$ of G is distance preserving. By symmetry considerations, we can assume that it is distance preserving from a vertex v_i on the left. It is easily seen that $\{v_1, v_1'\} \in \overline{E}$ and $\{v_k', v_{k+1}'\} \notin \overline{E}$. Let T' be the tree obtained from $\overline{T}$ by replacing $\{v_{k-1}', v_k'\}$ by $\{v_k', v_{k+1}'\}$. It can be checked that for $k \geq 3$ and $r \geq 6$, $D(T') < D(\overline{T})$, which contradicts the fact that $\overline{T}$ is a MAD spanning tree of G. This shows that for $k \geq 3$ and $r \geq 6$, the graph of Fig. 1(a) gives a negative answer to Question 1 [1].

Consider the graph G of Fig. 1(b). A spanning tree T of G is obtained by deleting an edge $\{v_i, v_j\}$ of the unique cycle of G. Such a tree is distance preserving from the vertex opposite to edge $\{v_i, v_j\}$ in G, i.e., from the vertex v_k such that $d_G(v_i, v_k) = d_G(v_j, v_k)$. On the other hand, the unique median of G is v_1. Let T_1 and T_2 denote respectively the DP spanning trees rooted at v_1 and v_2. It can be checked that for $k \geq 4$, $D(T_2) < D(T_1)$. This gives a negative answer to the second question [1].

3. A Simple Bound

Consider a graph G of order n that is homogeneous, i.e., all edges are of length one. We establish in this section, Proposition 3.1 below which is a weak version of the main result presented in this paper. This will give the reader a flavor of our approach.

Proposition 3.1. *In a MAD spanning tree of G, any branch $B' = (V', E')$ of size less than $\sqrt[3]{n}$, with root $\overline{r}$, is distance preserving from $\overline{r}$, with respect to the subgraph of G induced by V'.*

Proof. Consider a MAD spanning tree $\overline{T} = (\overline{V}, \overline{E})$ of G and a branch $B' = (V', E')$ with root $\overline{r}$ and size less than $\sqrt[3]{n}$. Assume that B' is not distance preserving from $\overline{r}$, with respect to the subgraph of G induced by V'. Clearly, $n \geq 4$. Let T^* be the spanning tree of G obtained from $\overline{T}$ by replacing B' by a spanning tree B^* of the subgraph of G induced by V', that is distance preserving from $\overline{r}$.

$$\begin{aligned} D(\overline{T}) &= D_{\overline{T}}(V', V') + D_{\overline{T}}(\overline{V} - V', \overline{V} - V') \\ &\quad + 2 \sum_{x \in V'} \sum_{y \notin V'} (d_{\overline{T}}(x, \overline{r}) + d_{\overline{T}}(\overline{r}, y)) \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} D(T^*) &= D_{T^*}(V', V') + D_{\overline{T}}(\overline{V} - V', \overline{V} - V') \\ &\quad + 2 \sum_{x \in V'} \sum_{y \notin V'} (d_{T^*}(x, \overline{r}) + d_{\overline{T}}(\overline{r}, y)). \end{aligned} \tag{3.2}$$

Hence

$$\begin{aligned} D(T^*) - D(\overline{T}) &= (D_{T^*}(V', V') - D_{\overline{T}}(V', V')) \\ &\quad + 2(n - |V'|) \left(\sum_{x \in V'} d_{T^*}(x, \overline{r}) - \sum_{x \in V'} d_{\overline{T}}(x, \overline{r}) \right). \end{aligned} \tag{3.3}$$

On the other hand, $D_{T^*}(V', V') \leq |V'|^3$ and $\sum_{x \in V'} d_{T^*}(x, \overline{r}) - \sum_{x \in V'} d_{\overline{T}}(x, \overline{r}) \leq -1$. It follows that $D(T^*) - D(\overline{T}) \leq |V'|^3 - 2(n - |V'|)$, with $|V'| \leq \sqrt[3]{n}$.

Consider the function $f : x \in [0, n] \mapsto f(x) = x^3 - 2(n - x)$. Its derivative is $f'(x) = 3x^2 + 2 > 0$, hence f is strictly increasing. Since $f(\sqrt[3]{n}) = -n + 2\sqrt[3]{n} < 0$ for $n \geq 4$, it follows that $f(x) < 0$ for $x < \sqrt[3]{n}$. As a consequence, $D(T^*) - D(\overline{T}) < 0$ for $|V'| < \sqrt[3]{n}$. This ends the proof of the proposition. □

In the forthcoming section, thanks to a deeper analysis, the bound $\sqrt[3]{n}$ of Proposition 3.1 is improved for graphs with integer edge lengths.

4. Total Distance Variation Following a Father-Change

Definition 4.1. Let T be a spanning tree of $G = (V, E, w)$ with root r. The father in T of a vertex u is the vertex v adjacent to u in the path from r to u in T. We denote by $T_r(v)$ the subtree of T rooted at v.

Definition 4.2. Let u be a vertex with father v in T, and let v' be a vertex different from v such that $\{u, v'\} \in E$ and $v' \notin T_r(u)$. The transformation that removes $\{u, v\}$ from T and replaces it by $\{u, v'\}$ is called a father-change.

A father-change is a particular case of the 1-move transformation extensively used in [3]. A 1-move transformation constructs a new spanning tree T' from T, by adding an edge $e = \{u, v\} \notin E(T)$ and removing an edge from the unique cycle created by e in T. Hereafter, we denote by B the branch of r containing v minus $T_r(u)$, and by B' the branch of r containing v' (see Fig. 2). It is important to note that if $v' = r$, then $B' = \emptyset$. From now, we denote $d = d_T(r, u)$, $d' = d_{T'}(r, u)$ and a is the number of vertices in the subtree $T_r(u)$.

4.1. *Father-change with branch change*

We suppose in this section that $B \neq B'$. We first note that if v and v' belong to different branches of r, then

$$\begin{aligned} D(T') - D(T) = {} & (D_{T'}(T_r(u), B) - D_T(T_r(u), B)) \\ & + (D_{T'}(T_r(u), B') - D_T(T_r(u), B')) \\ & + a(n - |B| - |B'| - a)(d' - d). \end{aligned} \tag{4.1}$$

Lemma 4.3. *Let $\Delta_B = D_{T'}(T_r(u), B) - D_T(T_r(u), B)$ be the variation of the sum of distances between the vertices of $T_r(u)$ and those of B as a result of the father-change. $\Delta_B \leq a|B|(d' + d - 2)$.*

Proof. Consider the path $Ch = (r, v_1, \ldots, v_k = v, u = v_{k+1})$ from r to u in T. Let $d_1, \ldots, d_k$ be the distances from r to $v_1, \ldots, v_k$ in T. It is easy to check that if $x \in T_r(u)$ and $y \in T_r(v_i)\backslash T_r(v_{i+1})$, then the variation of the distance from x to y is exactly $d' + d_i - (d - d_i) = d' + 2d_i - d \leq d' + 2d_k - d$. Since $B = \bigcup_{i=1}^{k} T_r(v_i)\backslash T_r(v_{i+1})$ it follows that $\Delta_B \leq a|B|(d' + 2d_k - d)$. Since $d_k \leq d - 1$, we deduce that $\Delta_B \leq a|B|(d' + d - 2)$. □

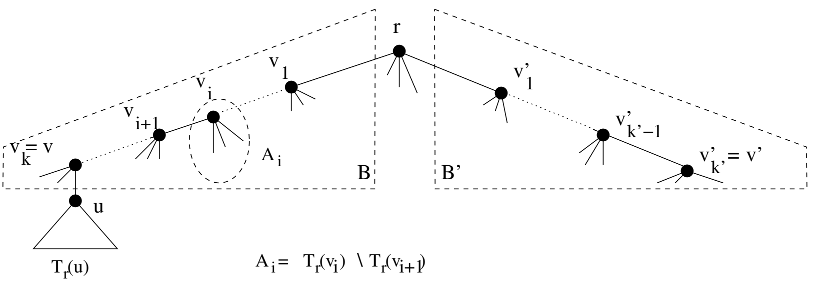

Fig. 2. Structure of T in the case of father-change with branch change.

Lemma 4.4. *Let $\Delta_{B'} = D_{T'}(T_r(u), B') - D_T(T_r(u), B')$ be the variation of the sum of distances between the vertices of $T_r(u)$ and those of B'. $\Delta_{B'} \le a|B'|(d' - d - 2)$.*

Proof. Suppose that B' is not empty. Consider the path $Ch' = (r, v'_1, \ldots, v'_{k'}, u)$ from r to u in T'. Let $d'_1, \ldots, d'_{k'}$ be the distances from r to $v'_1, \ldots, v'_{k'}$ in T'. It is easy to check that if $x \in T_r(u)$ and $y \in T_r(v'_i) \backslash T_r(v'_{i+1})$ for $i = 1, \ldots, k' - 1$ or $y \in T_r(v'_{k'})$ then the variation of the distance from x to y is $d' - d'_i - (d + d'_i) = d' - 2d'_i - d \le d' - 2d'_1 - d$. Hence $\Delta_{B'} \le a|B'|(d' - 2d'_1 - d)$. Since $d'_1 \ge 1$, we deduce $\Delta_{B'} \le a|B'|(d' - d - 2)$. If B' is empty, the father-change links u to the root r and the inequality holds since $|B'| = 0$. □

We will use Lemmas 4.3 and 4.4 to characterize the father-changes with branch change that decrease the total path length.

Proposition 4.5. *Let w^+ the maximum edge-length in G and let $\Delta = D(T') - D(T)$ the total variation of the sum of distances after a father-change with branch change such that $d' < d$. If $|B| + a \le \sqrt{\frac{n}{2w^+}}$ then $\Delta < 0$.*

Proof. From Eq. (4.1), $\Delta = \Delta_B + \Delta_{B'} + a(n - (|B| + a) - |B'|)(d' - d)$. Its follows from Lemmas 4.3 and 4.4 that

$$\begin{aligned}
\Delta &\le a|B|(d' + d - 2) + a|B'|(d' - d - 2) + a(n - a - |B| - |B'|)(d' - d) \\
&\le a|B|(d' + d - 2) + a|B'|((d' - d - 2) - (d' - d)) + a(n - a - |B|)(d' - d) \\
&\le a|B|(d' + d - 2) - 2a|B'| + a(n - a - |B|)(d' - d) \\
&\le a|B|(d' + d - 2) + a(n - a - |B|)(d' - d).
\end{aligned}$$

Let $g(a, d, d', |B|) = a|B|(d' + d - 2) + a(n - a - |B|)(d' - d)$. $\partial g(a, d, d', |B|)/\partial d' = a|B| + a(n - a - |B|) = a(n - a) > 0$. Since $d' \le d - 1$, we deduce that

$$\begin{aligned}
\Delta &\le g(a, d, d', |B|) \\
&\le g(a, d, d - 1, |B|) \\
&\le a|B|(2d - 3) - a(n - a - |B|) \\
&\le a(|B|(2d - 3) - n + |B| + a).
\end{aligned}$$

There are two cases:

(1) If $|B| + a \le 2d - 3$, then $\Delta \le a[|B|(2d - 3) - n + (2d - 3)] = a[(|B| + 1)(2d - 3) - n]$. Since $d \le w^+(|B| + 1)$, it follows that $\Delta \le a((|B| + 1)(2w^+(|B| + 1) - 3) - n) = a(2w^+(|B| + 1)^2 - 3(|B| + 1) - n)$. Since $2w^+x^2 - 3x - n < 0$ for $0 \le x \le \sqrt{\frac{n}{2w^+}}$, it follows that $\Delta < 0$ for $|B| + 1 \le \sqrt{\frac{n}{2w^+}}$.
(2) If $|B| + a > 2d - 3$, then $\Delta < a[|B|(|B| + a) - n + |B| + a] = a[(|B| + a)^2 - n + (1 - a)(|B| + a)]$. Since $1 \le a$, it follows that $\Delta < a[(|B| + a)^2 - n]$. If $|B| + a \le \sqrt{n}$, we derive $\Delta < 0$.

As a consequence, if $|B| + a \le \sqrt{\frac{n}{2w^+}}$ then $\Delta < 0$. □

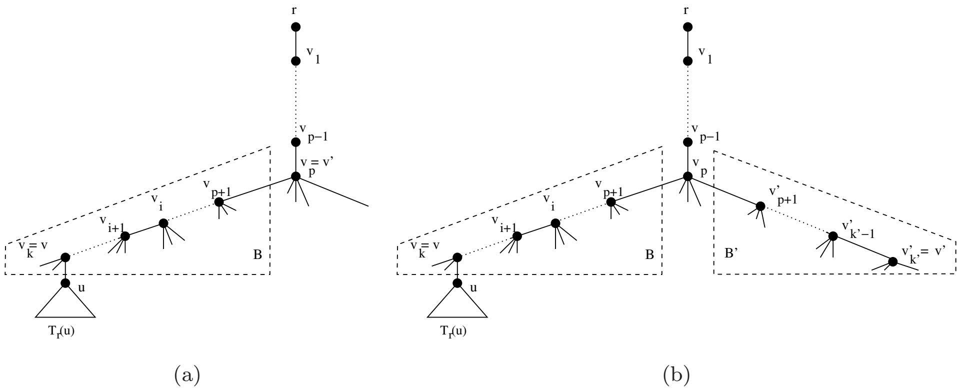

Fig. 3. Father-change in the same branch.

4.2. *Father-change in the same branch*

Let B_{Ch} be the branch of r in which the change occurs. We will show in this section that Proposition 4.5 also holds in the case of a father-change in the same branch. Consider the path $Ch = (r, v_1, \ldots, v_k = v, u = v_{k+1})$ from r to u in T. According to the position of v' in B_{Ch}, there are two cases:

Case 1: v' is on the path from r to u (see Fig. 3(a));

Case 2: v' does not belong to the path from r to u (see Fig. 3(b)).

In both cases, we can apply Proposition 4.5 by letting v_p play the role previously assumed by vertex r.

We are now able to give the main result of this paper.

Theorem 4.6. *Let $\overline{T} = (\overline{V}, \overline{E})$ be a MAD spanning tree of an undirected graph $G = (V, E, w)$ with integer lengths. Let $T' = (V', E')$ be a subtree of $\overline{T}$ consisting of a vertex $\overline{r}$ and the union of branches of $\overline{r}$ that are each of size less than or equal to $\sqrt{\frac{n}{2w^+}}$. T' is distance preserving from $\overline{r}$ in the subgraph of G induced by V'.*

Proof. We proceed by contradiction. So, let G' be the subgraph of G induced by V' and let u be a vertex of T' such that $d_{T'}(\overline{r}, u) > d_{G'}(\overline{r}, u)$, and let $v_0 = \overline{r}, v_1, \ldots, v_k = u$ be a shortest path of G'. Let i be the minimum index such that $\{v_i, v_{i+1}\} \notin E'$. Let us denote by b the length of the path $v_0, \ldots, v_i$ in $\overline{T}$, by l the length of the path $v_0, \ldots, v', v_{i+1}$ in $\overline{T}$ and by c the length of the edge $\{v_i, v_{i+1}\}$ (see Fig. 4). By definition of a shortest path, $b + c \leq l$. Moreover, if $b + c < l$, then the father-change applied to v_{i+1} by replacing in $\overline{T}$ $\{v_{i+1}, v'\}$ by $\{v_{i+1}, v_i\}$ leads from Proposition 4.5 to a spanning tree T^* such that $D(T^*) < D(\overline{T})$, and this contradicts the optimality of $\overline{T}$. As a consequence $b + c = l$. This leads to a new shortest path $v'_0 = \overline{r}, v'_1, \ldots, v'_j = v_{i+1}, v_{i+2}, \ldots, v_k = u$ of G' where the subpath $v'_0 = \overline{r}, v'_1, \ldots, v'_j = v_{i+1}$ is a path of $\overline{T}$. In this new path, the first edge that does not

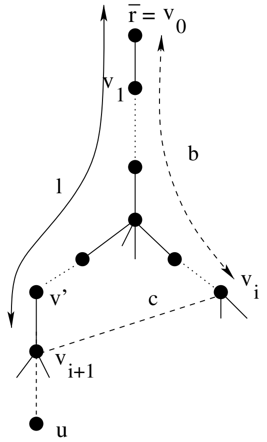

Fig. 4. The initial shortest path is in dotted line.

belong to $\overline{E}$ is of the form $\{v_j, v_{j+1}\}$ with $j > i$. Clearly we can apply to $\{v_j, v_{j+1}\}$ the analysis carried out for $\{v_i, v_{i+1}\}$. The successive application of this sequence pushes to the right the first edge that does not belong to $\overline{E}$. After at most k steps we will have a shortest path from $\overline{r}$ to u where all edges belong to $\overline{E}$. This contradicts the fact that $d_{T'}(\overline{r}, u) > d_{G'}(\overline{r}, u)$. □

Theorem 4.6 gives a necessary condition for $\overline{T}$ to be a MAD spanning tree of a graph G of order n. Distance preserving spanning trees from a root $\overline{r}$ are much easier to enumerate than all spanning trees. Indeed, starting from the root, one must introduce nodes at distance 1, then those at distance 2 etc. Moreover when a node at distance k is introduced, one needs to consider only edges that link it to nodes at distance $k-1$. Since the condition of Theorem 4.6 deals with branches that must be distance preserving MAD spanning trees, it can be checked by an enumerative approach for small values of $|V'|$.

5. Conclusion

We have shown in this paper that in a MAD spanning tree of a graph G whose edges are of positive integer lengths, a subtree $H' = (V', E')$ consisting of a vertex $\overline{r}$ and the union of branches of $\overline{r}$ that are each of size less than or equal to $\sqrt{\frac{n}{2w^+}}$, where w^+ is the maximum edge-length in G, is a distance preserving spanning tree of the subgraph of G induced by V'. As the results of Wong [4] and Dankelmann [1], this property is a new contribution to the understanding of the relations between MAD and distance preserving spanning trees. It is also a necessary condition that may be useful in practice because it can be checked for reasonable values of $|V'|$ on candidate MAD spanning trees.

References

[1] P. Dankelmann, A note on MAD spanning trees, *J. Combin. Math. Combin. Comput.* **32** (2000) 93–95.

[2] R. C. Entringer, Distance in graphs: trees, *J. Combin. Math. Combin. Comput.* **24** (1997) 65–84.

[3] M. Tchuente, P. Y. Melatagia, J. M. Nlong II and Y. Denneulin, On the minimum average distance spanning tree of the hypercube, *Acta Applicandae Mathematicae* **102**(3) (2008) 219–236.

[4] R. T. Wong, Worst-case analysis of network design problem heuristics, *SIAM J. Algebraic Discrete Methods* **1**(1) (1980) 51–63.

[5] B. Y. Wu, K. M. Chao and C. Y. Tang, Approximation algorithms for the shortest total path length spanning tree problem, *Discrete Appl. Math.* **105** (2000) 273–289.

[6] B. Y. Wu, G. Lancia, V. Bafna, K. M. Chao, R. Ravi and C. Y. Tang, A polynomial-time approximation scheme for minimum routing cost spanning trees, *SIAM J. Comput.* **29**(3) (1999) 761–778.