Distance in Graphs: Trees

Roger Entringer
Department of Mathematics and Statistics
University of New Mexico
Albuquerque, New Mexico 87131, USA
entring@math.unm.edu

ABSTRACT. The *distance of a vertex* u in a connected graph G is defined by $\sigma_G(u) := \sum_{v\in V(G)} d(u,v)$ and the *distance of* G is given by $\sigma(G) := \frac{1}{2}\sum_{u\in V(G)} \sigma(u)$ $(= \sum_{\{u,v\}\subseteq V(G)} d(u,v)$. Thus the average distance between vertices in a connected graph G of order n is $\sigma(G)/\binom{n}{2}$. These graph invariants have been studied for the past fifty years. Here we discuss some known properties and present a few new results together with several open problems. We focus on trees.

1 Introduction

Fifty years ago Wiener [23], [24], [25], [26], and [27] presented an empirical formula for predicting the boiling points of certain hydrocarbons; a formula that takes into account the molecular bond structure. It was known that isomers, i.e., compounds with the same chemical formula but with different chemical structure, could have different boiling points. In Figure 1 two such isomers, both with the formula C_7, H_{16}, are pictured. Note that only the carbon atoms are shown; since they each have valence 4 while carbon has valence 1, it is clear how many hydrogen atoms must be attached to each carbon atom.

To explain the different boiling points of various saturated paraffins (including different boiling points of isomers) it can be argued that compounds with a less "compact" molecular structure would boil at higher temperatures since they were subject to more entanglement during motion. This led to Wiener's use of distance as a measure of molecular "compactness" and so distance was incorporated as an additive term in his empirical formula for boiling points.

JCMCC 24 (1997), pp. 65-84

The boiling points of the paraffins (see [23]) are closely approximated by the formula $t_B = aw + bp + c$ where a, b, and c are constants for a given isometric group and p and w are structural variables defined as follows. The polarity number p is defined as the number of pairs of carbon atoms which are separated by three carbon-carbon bonds and w is the sum of the distances between carbon atoms in the molecule. Thus the heptane 2,3-dimethylpentane (if all hydrogen atoms are deleted the resulting structure is represented by the unique tree of order 7 having two adjacent vertices of degree 3) has $p = 6$ and $w = 46$. Wiener has shown (see references above) that similar results hold for molecular refraction, molecular volume, heats of formation, and vapor pressure.

Much of the material contained in this limited survey that has not appeared elsewhere is available in detailed form from the author. Our goal here is to present a coherent overview of distance vis-à-vis trees.

2 Background

The *distance of a vertex* u in a connected graph G is defined by

$$\sigma_G(u) := \sum_{v \in V(G)} d(u, v).$$

We will suppress the subscripts on σ when no confusion can arise.

In Figures 1 and 2 we show the distance of each of the vertices in the respective graphs. In both figures vertices with minimum distance have been darkened.

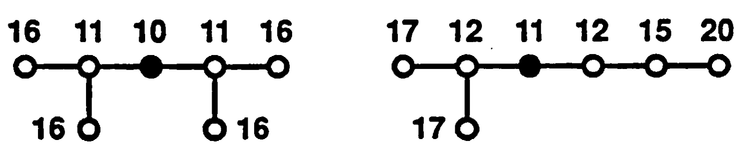

Figure 1. Two isomers of Heptane

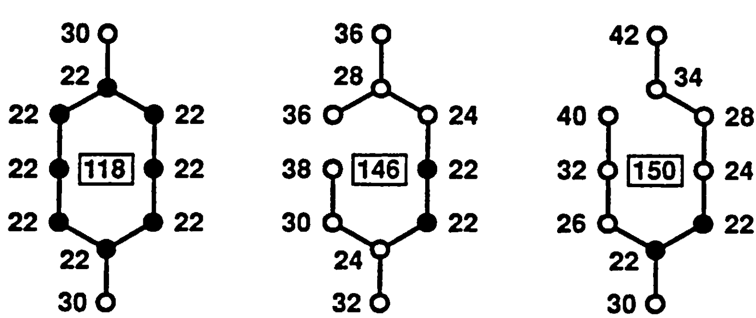

Figure 2. An unlabeled graph and its two spanning trees

Various techniques for the rapid calculation of distances of vertices are available; the following is one such.

Observation A. (Entringer, Jackson, and Snyder [5]) *Suppose a and b are adjacent vertices of a connected graph G of order n. Let A be the set of vertices closer to a than to b and let B be the set of vertices closer to b than to a. Then $\sigma(a) - \sigma(b) = |B| - |A|$.*

Noting that $|B| + |A| = n$ we have the convenient formula $\sigma(b) = \sigma(a) + n - 2|B|$. Thus, after using the definition to determine the distance of one vertex, only simple calculations are needed to iteratively determine the distances of all remaining vertices.

The *distance of a connected graph* G is defined by

$$\sigma(G) := \sum_{\{u,v\}\subseteq V(G)} d(u,v) = \frac{1}{2} \sum_{u\in V(G)} \sigma(u).$$

Thus the left tree of Figure 1 has distance 48 and the right tree has distance 52. In Figure 2 the distances of the trees are the numbers enclosed in boxes.

Canfield, Robinson, and Rouvray [4] have described a technique for computing the distance of trees that depends on recursively removing a root vertex and determining the distance of the subtrees remaining. A further method is discussed and exemplified by Gutman, Yeh, Lee, and Chen [11]. Also, in our concluding remarks we discuss another technique used by Wiener [23] for determining the distance of trees.

The *average distance of a connected graph* G of order n is given by

$$\mu(G) := \sigma(G)/\binom{n}{2}.$$

Thus the average distance of G is just that: the mean distance between vertices.

In Table 1 we present the distance of all vertices, the distance of G, and the average distance of G for selected graphs G. We require vertex labelings for those graphs that are not vertex transitive.

The vertices of the path P_n are labeled sequentially v_1 to v_n from one end vertex to the other. Thus

$$\sigma(v_i) = \sum_{j=1}^{j=n} |j - i| = \binom{n+1-i}{2} + \binom{i}{2} = \binom{n+1}{2} + 2\binom{i}{2} - ni.$$

The center of $K_{1,n-1}$ is labeled v_1. In the partition of $V(K_{r,s})$ into parts R and S with $|R| = r$ and $|S| = s$, we label the vertices in R from 1 to r and the vertices in S from $r+1$ to $r+s$.

If v is any vertex of the hypercube Q_n then

$$\sigma(v) = \sum_{i=1}^{n} i\binom{n}{i} = n2^{n-1}.$$

Graph	$\sigma(v_i)$	$\sigma(G)$	$\mu(G)$
P_n	$\binom{n+1}{2} + 2\binom{i}{2} - ni$	$\binom{n+1}{3}$	$\frac{n+1}{3}$
$K_{1,n-1}$	$n-1,\ i=1$ $2n-3,\ 2 \le i \le n$	$(n-1)^2$	$2 - \frac{2}{n}$
C_n	$\lfloor \frac{n^2}{4} \rfloor$	$\frac{n}{2} \lfloor \frac{n^2}{4} \rfloor$	$\frac{1}{n-1} \lfloor \frac{n^2}{4} \rfloor$
K_n	$n-1$	$\binom{n}{2}$	1
$K_{r,s}$	$s+2(r-1),\ 1 \le i \le r$ $r+2(s-1),\ r+1 \le i \le r+s$	$rs+r^2+s^2$ $-r-s$	$2 - \frac{2rs}{(r+s)(r+s-1)}$
Q_n	$n2^{n-1}$	$n4^{n-1}$	$\frac{n}{2} + \frac{n}{2^{n+1}-2}$

Table 1. Distances in some special graphs

We note for future reference that the star and the path attain the extreme values for distance of trees.

Theorem B. (Entringer, Jackson, and Snyder [5]) *If T is a tree of order n then*

$$(n-1)^2 \le \sigma(T) \le \binom{n+1}{3}.$$

The lower bound is realized only by $K_{1,n-1}$ and the upper only by P_n.

One can compare these extremal values with the expected values for certain classes of trees.

Theorem C. (Entringer, Meir, Moon, and Székely [6]) *The expected value of $\sigma(T)$ over all (ordered) (rooted labeled) (or rooted binary) trees of order n is asymptotic to $Cn^{5/2}$ where C is a constant dependent only on the class of trees.*

As we will see, the centroid of a tree plays an important role in the study of distance. Let us recall the relevant definitions.

A maximal subtree containing a vertex v of a tree T as an end vertex will be called a *branch* of T at v. Figure 3 shows a tree with three branches at the vertex v.

The *weight* of a branch B, denoted by $bw(B)$, is the number of edges in it. The *branch weight* of a vertex v is the maximum of the weights of the branches at v. In Figure 3 the branches at v have weights 1, 2, and 3. The branch weights of each of the vertices are marked in Figure 4 where the vertices with minimum weight have been darkened. The *centroid* of a tree T, denoted by $C(T)$, is the set of vertices v of T with minimum branch weight. The centroid of the tree in Figure 3 is $\{v\}$ while each of the trees of Figure 4 has a doubleton centroid.

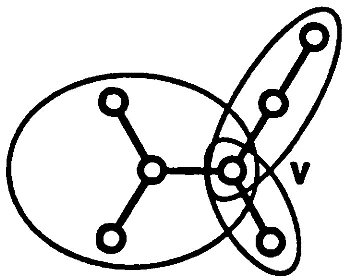

Figure 3. The branches of a tree at a vertex

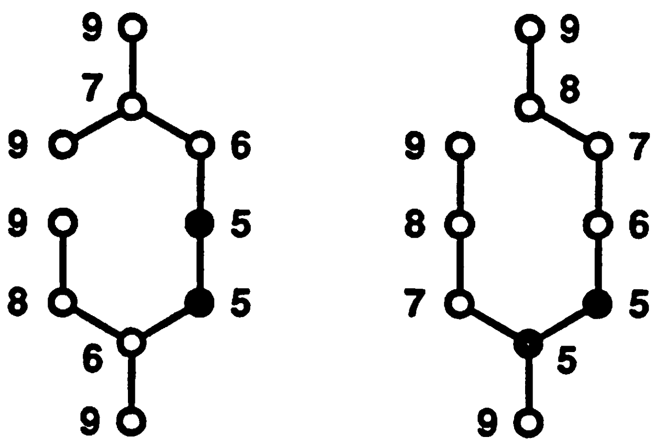

Figure 4. Branch weights of vertices

Long ago Jordan discovered an important characterizing property of centroid vertices.

Theorem D. (Jordan [17]) *If $C = C(T)$ is the centroid of a tree T of order n then one of the following holds:*

(i) $C =: \{c\}$ *and* $bw(c) \leq (n-1)/2$,

(ii) $C =: \{c_1, c_2\}$ *and* $bw(c_1) = bw(c_2) = n/2$.

In both cases, if $v \in V(T)\backslash C$ then $bw(v) > n/2$.

A comparison of Figures 2 and 4 gives an instance of the following important result.

Theorem E. (Zelinka [30]) *The set of vertices with minimum distance in a tree T is the centroid of T.*

The following result arose in the study of distributed processing in networks. Using it Gerstel and Zaks showed that, given a network of a tree topology, choosing a centroid vertex and then routing all the information through it is the best possible strategy, in terms of worst-case number of messages sent during any execution of any distributed sorting algorithm.

Theorem F. (Gerstel and Zaks [9]) *If T is a tree of order n then $v \in C(T)$ iff $V(T)$ ($V(T)\backslash v$ if n is odd) con be partitioned into pairs $\{u_i, v_i\}$ such that each $u_i - v_i$ path contains v. Furthermore, $\sum_{i=1}^{\lfloor n/2 \rfloor} d(u_i, v_i)$ is largest among all such partitions.*

If we modify the proof of Gerstel and Zaks appropriately we obtain one that in summary form reads as follows.

We note that, by the concluding statement in Jordan's theorem, if v is not a centroid vertex of the tree T then by the pigeonhole principle the desired portion into pairs cannot exist. On the other hand, if $v \in C(T)$ and two end vertices u and w are chosen, one from each of the two branches of T at v with greatest weight, then $V(T)$ has the desired partition into pairs. One of the pairs is $\{u, w\}$ and the remaining can be obtained from $T - \{u, w\}$ by induction since it follows from Jordan's theorem that $C(T - \{u, w\}) = C(T)$. Furthermore, if $v \in C(T)$ and $\{w_i, w_i'\}$ is any pairing of the vertices of $V(T)$ ($V(T)\backslash v$ if n is odd) then

$$\sum_{i \geq 1} d(w_i, w_i') \leq \sum_{i \geq 1} [d(w_i, v) + d(v, w_i')] = \sigma(v).$$

so that the sum of the distances between members of pairs is maximized iff the pairing is one of the desired partitions into pairs.

3 Average distance and inverse degree

Definition. Let G be a graph with no isolated vertices. The *inverse degree* of G is given by

$$\mathrm{ID}(G) := \sum_{v \in V(G)} \frac{1}{d(v)}.$$

Among the many conjectures produced by the computer program *Graffiti* devised by Fajtlowicz is the following; it was proved by Shi.

Conjecture. (Fajtlowicz: Graffiti #592); **Theorem G.** (R. Shi [21]) *For a tree of order $n \geq 2$, $\mu(T) \cdot \mathrm{ID}(T) \geq n$ and this bound is sharp.*

We will use the following to construct a short proof of Shi's result.

Theorem 1. *If T is a tree of order $n \geq 2$ then*

$$\frac{n+2}{2} \leq ID(T) \leq \frac{n^2-2n+2}{n-1}.$$

The lower bound is achieved iff $T = P_n$ and the upper iff $T = K_{1,n-1}$.

Proof: Let T be a tree of order $n \geq 2$ for which $\mathrm{ID}(T)$ is minimum. To show that $T = P_n$ we suppose, to the contrary, that $v \in V(T)$ with $d(v) \geq 3$. Choose vertices u and w adjacent to v and let x be an end vertex, different from v, of the branch of T at v containing u. Set $T' = T - vw + xw$. Then

$$\mathrm{ID}(T) - \mathrm{ID}(T') = 1 - \frac{1}{2} + \frac{1}{d(v)} - \frac{1}{d(v)-1} = \frac{1}{2} - \frac{1}{d(v)(d(v)-1)} > 0.$$

Since this is impossible, $T = P_n$.

Now let T be a tree of order $n \geq 2$ for which $\mathrm{ID}(T)$ is maximum. To show that $T = K_{1,n-1}$ suppose, to the contrary, that T contains two vertices v and w with $d(v) = \Delta(T)$ and $d(w) > 1$. Choose u adjacent to w so that w lies in the $u-v$ path in T. Set $T' = T - uw + uv$. Then

$$\begin{aligned}\mathrm{ID}(T') - \mathrm{ID}(T) &= \frac{1}{d(v)+1} - \frac{1}{d(v)} + \frac{1}{d(w)-1} - \frac{1}{d(w)} \\ &= \frac{1}{d(w)(d(w)-1)} - \frac{1}{d(v)(d(v)+1)} > 0.\end{aligned}$$

Since this is impossible, $T = K_{1,n-1}$.

The following result is an immediate consequence of Theorem B and Theorem 1.

Theorem 2. *If T_1 and T_2 are trees of order $n \geq 2$ then*

$$\mu(T_1) \cdot ID(T_2) \geq n + 1 - \frac{2}{n}.$$

Although this is only a mild improvement over Theorem G, the fact that $\sigma(T)$ is minimized by $K_{1,n-1}$ while $\mathrm{ID}(T)$ is minimized by P_n suggests that considerable improvement may be obtained using analogs of Theorems B and 1 with maximum degree $\Delta(T)$ prescribed. This is pursued in Reference [8].

4 Some extremal ratios

The following two classes of graphs appear frequently as extremal trees in the study of distance. The first class, $\mathcal{T}(n, \Delta)$, consists of trees with root

v and maximum degree Δ. Furthermore, if n is the order of such a tree we write $n = 1 + \Delta\frac{(\Delta-1)^r-1}{\Delta-2} + s$ where $0 \leq s < \Delta(\Delta-1)^r$ and require v to have exactly $\Delta(\Delta-1)^{i-1}$ vertices distance i from v for $1 \leq i \leq r$ and s vertices distance $r+1$ from v. Such a tree is well-defined except for the neighborhoods of the end vertices.

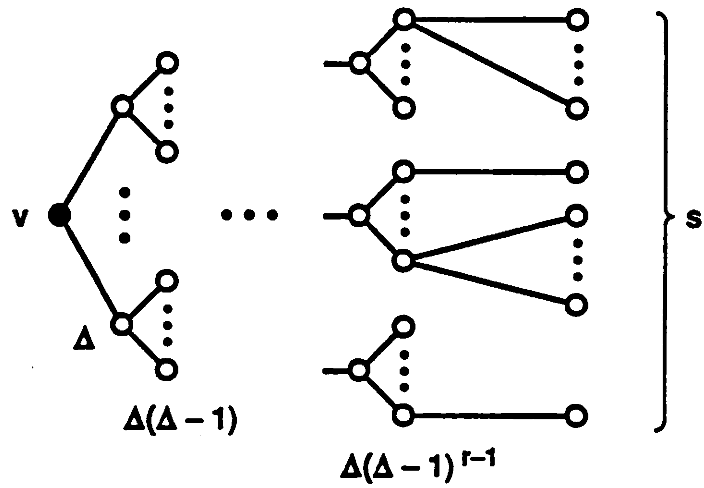

Figure 5. The class $\mathcal{T}(n, \Delta)$ of trees

The second class, $T(n, r)$, consists of a path of length $r-1$ together with $n-r$ additional vertices all adjacent to the same end vertex of the path.

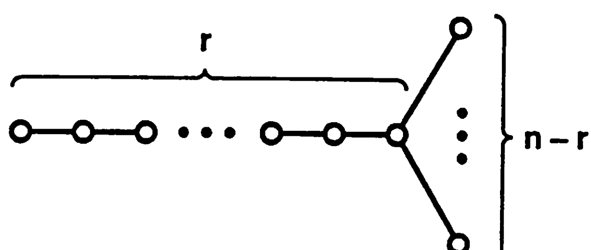

Figure 6. The tree $T(n, r)$

Before stating, without proofs, some results involving extremal ratios (the proofs are lengthy) we state and prove an easier result to show the type of arguments used. If T is a tree of order n, with maximum degree $\Delta = 2$, and $v \in C(T)$, then obviously $T = P_n$ and an easy calculation gives $\sigma_T(v) = \lfloor n^2/4 \rfloor$. For $\Delta > 2$ we have the following.

Theorem 3. *If T is a tree with order n, maximum degree $\Delta > 2$, $v \in C(T)$, and the integers r and s are defined by $n =: 1 + \Delta\frac{(\Delta-1)^r-1}{\Delta-2} + s$,*

$0 \le s < \Delta(\Delta-1)^r$, *then*

$$s(r+1) + \frac{\Delta}{(\Delta-2)^2}\{[(\Delta-2)(r+1)-1](\Delta-1)^{r+1}+1\}$$
$$\le \sigma(v) \le \begin{cases} \left\lfloor \frac{n^2}{4} \right\rfloor - \frac{(\Delta-1)(\Delta-2)}{2}, & \Delta \le \frac{n}{2} \\ \binom{n-\Delta+1}{2} + \Delta - 1, & \Delta \ge \frac{n}{2} \end{cases}.$$

The lower bound is achieved iff $T \in \mathcal{T}(n,\Delta)$ and the upper bound is achieved iff $T = T(n, n-\Delta+1)$.

Proof: Let T be a tree of order n and maximum degree Δ and suppose $v \in C(T)$. Set n_i equal to the number of vertices of T whose distance from v is i. Then

$$n = \sum_{i\ge 0} n_i \quad \text{and} \quad \sigma(v) = \sum_{i \ge 0} i n_i.$$

Since $n_1 \le \Delta$ and $n_i \le (\Delta-1)n_{i-1}$ for $i \ge 2$, we have

$$\sigma(v) \le \sum_{i=1}^{r} i\Delta(\Delta-1)^{i-1} + s(r+1) = \Delta\frac{\partial}{\partial\Delta}\sum_{i=0}^{r}(\Delta-1)^i + s(r+1)$$
$$= s(r+1) + \frac{\Delta}{(\Delta-2)^2}\{[(\Delta-2)(r+1)-1](\Delta-1)^{(r+1)}+1\}$$

with equality holding iff $n_1 = \Delta$ and $n_i = (\Delta-1)n_{i-1}$, for $2 \le i \le r$, i.e., $T \in \mathcal{T}(n,\Delta)$.

Now suppose T is a tree of order n with maximum degree Δ, having $v \in C(T)$ and of all such trees T is chosen so that $\sigma_T(v)$ is maximum. We prove $T = T(n, n-\Delta+1)$ by establishing a series of claims. Choose vertex w to satisfy $d_T(w) = \Delta$. The contrapositive of the final statement of Jordan's centroid characterization is used repeatedly in the proof of the following.

Claim 1. *If $v = w$ every branch of T at w is a path.*

Suppose, to the contrary, that B is a branch of T at w containing a vertex x with $d_T(x) > 2$. Let z $(\ne w)$ be an end vertex of B chosen so that x lies in the $w-z$ path and let y a vertex adjacent to x and not on the $w-z$ path in T. Denote the weight of the branch of T at x containing y by a. Define the tree $T' := T - xy + yz$. Then T' has maximum degree Δ, $v \in C(T')$, but $\sigma_{T'}(v) - \sigma_T(v) = a[d(v,z) - d(v,x)] > 0$. Since this is impossible, the claim follows.

The following claim can be proved in the same manner.

Claim 2. *Every branch of T at w not containing v is a path and every branch of T at v not containing w is a path.*

Now suppose $v \ne w$ and let P be the $v-w$ path in T.

Claim 3. *If $x \in V(P)\backslash\{v, w\}$ then $d_T(x) = 2$.*

Suppose, to the contrary, that $x \in V(P)\backslash\{v, w\}$ and x is adjacent to a vertex y not in $V(P)$. Let $z \neq w$ be an end vertex of some branch of T at w not containing v. Define $T' := T - xy + yz$. Then T' has maximum degree Δ, $v \in C(T')$, but $\sigma_{T'}(v) > \sigma_T(v)$ which is impossible.

Claim 4. *If $v \neq w$ then $d_T(v) = 2$.*

Suppose, to the contrary, that there are two branches A and B of T at v neither of which contains w. Let $x \neq v$ and $y \neq v$ be end vertices of A and B, respectively, and label the vertex of A adjacent to x as z. We may assume $d(v, x) \leq d(v, y)$.

If $bw(B) < \lfloor n/2 \rfloor$ define $T' := T - xz + xy$. Then T' has maximum degree Δ, $v \in C(T')$, but $\sigma_{T'}(v) - \sigma_T(v) = d(v, y) - d(v, z) > 0$, which is impossible.

If $bw(B) = \lfloor n/2 \rfloor$ label the vertex of A adjacent to v as u and choose an end vertex $t \neq w$ of T so that w lies in the $x - t$ path of T. Define the tree $T' = T - vu + ut$. Then T' has maximum degree Δ, $v \in C(T')$, but $\sigma_{T'}(v) - \sigma_T(v) = d(v, t)bw(A) > 0$ which is impossible.

The next claim is simply a summary of the previous four.

Claim 5. *Every branch of T at w is a path.*

It remains to show that at most one branch of T at w has weight greater than 1. Let us suppose otherwise and consider the following two cases.

Case 1. $v \neq w$. Denote by w_i, $1 \leq i \leq \Delta - 1$ the vertices adjacent to w and not in the branch of T at w containing v. Suppose, for some i, $1 \leq i \leq \Delta - 1$, $d(w_i) > 1$. We may suppose $i = 1$. Set $T' := T - \{ww_i \mid 2 \leq i \leq \Delta - 1\} + \{w_1w_i \mid 2 \leq i \leq \Delta - 1\}$. Then T' has maximum degree Δ (at w_1), $v \in C(T')$, but $\sigma_{T'}(v) - \sigma_T(v) \geq \Delta - 2 > 0$, which is impossible.

Case 2. $v = w$. There are two subcases to consider.

If some branch B of T at w has weight $\lfloor n/2 \rfloor$ choose a vertex w_1 with degree 2 and adjacent to w but not in B. Label the remaining vertices adjacent to w and not in B as w_i, $2 \leq w \leq \Delta - 2$. As above set $T' := T - \{ww_i \mid 2 \leq i \leq \Delta - 1\} + \{w_1w_i \mid 2 \leq i \leq \Delta - 1\}$. Then T' has maximum degree Δ (at w_1), $v \in C(T')$, but $\sigma_{T'}(v) - \sigma_T(v) \geq \Delta - 2 > 0$, which is impossible.

If every branch of T at w has weight less than $\lfloor n/2 \rfloor$ let A and B be two branches of T at w with branch weight at least 2. Choose end vertices x ($\neq w$) and y ($\neq w$) from A and B. We may assume $d(w, x) \leq d(w, y)$. Let z be the vertex adjacent to x in T. Set $T' = T - xz + xy$. Then T' has maximum degree Δ, $v \in C(T')$, but $\sigma_{T'}(v) - \sigma_T(v) = d(w, y) - d(w, z) > 0$, which is impossible.

We conclude that $T = T(n, n - \Delta + 1)$; easy calculations complete the proof of the theorem.

If we optimize the above extremal functions with respect to Δ we obtain the following specialization.

Corollary 4. *If T is a tree with order $n \geq 2$ and $v \in C(T)$, then*

$$n-1 \leq \sigma(v) \leq \left\lfloor \frac{n^2}{4} \right\rfloor$$

with the equalities holding iff $T = K_{1,n-1}$ or $T = P_n$, respectively.

Theorem H. (Barefoot, Entringer, and Szèkely [1]) *If w and u are end vertices of the tree T of order $n \geq 2$ and the integers $k \geq 1$ and s are defined by $2n = k^2 + s$, $0 \leq s \leq 2k$, then*

$$\frac{\sigma_T(w)}{\sigma_T(u)} \leq 2\frac{(n-)r + 2(n-1)}{r^2 - 3r + 4(n-1)} - 1$$

where

$$r = \begin{cases} \lfloor 2\sqrt{n} \rfloor - 2, & 0 \leq s \leq k-6 \\ \lfloor 2\sqrt{n} \rfloor - 1, & k-5 \leq s \leq 2k \end{cases}.$$

For $n \geq 5$ equality is achieved iff $T = T(n,r)$.

Theorem I. (Barefoot, Entringer, and Szèkely [1]) *If T is a tree of order $n \geq 2$, w is an end vertex of T, $v \in C(T)$, and $k \geq 1$ and s are defined by $2n = k^2 + s$, $0 \leq s \leq 2k$, then*

$$1 + 4\frac{n-2}{n^2 - 2n + a} \leq \frac{\sigma_T(w)}{\sigma_T(v)} \leq \frac{2rn - r^2 - r}{r^2 + 2n - 3r}$$

where $a = 8$ if n is even, $a = 5$ if n is odd, and

$$r = \begin{cases} \lfloor \sqrt{2n} \rfloor - 1, & 0 \leq s \leq k-4 \\ \lfloor \sqrt{2n} \rfloor, & k-3 \leq s \leq 2k \end{cases}.$$

For $n \geq 3$ the lower bound is obtained iff T consists of a path $u_1 u_2 \ldots u_{n-1}$ together with an additional vertex w adjacent to $u_{\lfloor (n-1)/2 \rfloor}$. The upper bound is obtained iff $T = T(n,r)$.

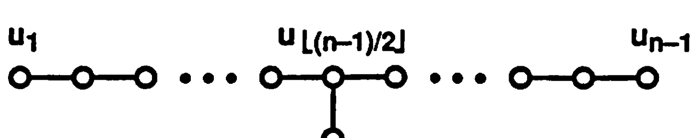

Figure 7. The unique graph achieving the lower bound

Theorem J. (Barefoot, Entringer, and Szèkely [1]) *If the tree T has order n and $v \in C(T)$ then*

$$\frac{1}{3}\left[2r+n-5+\frac{2n(n-1)(r-3)}{r^2-r(2n-1)}\right] \le \frac{\sigma(T)}{\sigma_T(v)} \le n-1.$$

The lower bound is achieved iff $T = T(n,r)$ with

$$r = \begin{cases} \lfloor\sqrt{2n}\rfloor, & 2n = k^2+s, 0 \le s \le \lfloor k/3 \rfloor \\ \lceil\sqrt{2n}\rceil, & 2n = k^2+s, \lfloor k/3\rfloor < s \le 2k \end{cases}.$$

and the upper bound is achieved iff $T = K_{1,n-1}$.

5 Minimum distance spanning trees

A spanning tree T of a connected graph G is said to be a *minimum distance spanning tree* if $\sigma(T)$ is minimum over all spanning trees of G. Solutions to the following problem are of great practical importance in the design of economical networks.

Problem. *Given a connected graph G, find a minimum distance spanning tree of G.*

Because Johnson, Lenstra, and Rinnooy-Kan [16] have shown that the problem is NP-complete, finding nearly optimal spanning trees is of interest. If T is any spanning tree of the connected graph G then it is obvious that $\sigma(T) \ge \sigma(G)$. On the other hand, we have the following result.

Theorem K. (Entringer, Kleitman, and Szèkely [7]) *A connected graph G of order n contains a spanning tree T satisfying $\sigma(T) \le 2\left(1-\frac{1}{n}\right)\sigma(G)$. Equality is achieved iff $G = K_n$ and $T = K_{1,n-1}$.*

Reference [7] contains two proofs; the first obtains from an averaging argument with a certain average being taken over a set of rooted spanning trees each of which is distance preserving from its root. A spanning tree T of a connected graph G is called *distance preserving* iff there is a vertex $v \in V(G)$ such that $d_T(u,v) = d_G(u,v)$ for all $u \in V(G)$. We will refer to v as a *root* of T and say T is *distance preserving* from v. For example, every breadth-first search tree is distance preserving from its root.

The second proof is so short we give it in its entirety. Choose a vertex r for which $\sigma_G(r)$ is minimum and let T be a rooted distance preserving spanning tree of G with root r. Choose $v \in C(T)$; by Theorem J above and Zelinka's result, $\sigma(T) \le (n-1)\sigma_T(v) \le (n-1)\sigma_T(r)$. Thus

$$\sigma(G) = \frac{1}{2}\sum_{u\in V(G)} \sigma_G(u) \ge \frac{n}{2}\sigma_G(r) = \frac{n}{2}\sigma_T(r) \ge \frac{n}{2(n-1)}\sigma(T).$$

We mentioned the two proofs since they lead to distinct algorithms for finding a spanning tree satisfying the inequality of the theorem.

Algorithm 1. *For every vertex v of G, construct a distance preserving spanning tree T_v and compute $\sigma(T_v)$. Select the tree which yields the smallest value.*

Algorithm 2. *Compute $\sigma_G(v)$ for each $v \in V(G)$ and then construct a distance preserving spanning tree T_v rooted at a vertex v for which $\sigma(v)$ is minimum.*

The left graph of Figure 2 has only the two trees pictured as distance preserving spanning trees. The center tree is distance preserving from the vertices of degree 3 and from each vertex adjacent to a vertex of degree 3 (but not simultaneously). The right tree is distance preserving from the vertices halfway between the vertices of degree 3 (again, not simultaneously). Thus the center tree would be selected by Algorithm 1 but either tree could be selected by Algorithm 2.

We can ask the following questions in an effort to focus on exploration of the relationships between the various concepts just discussed.

Question 1. *Does every graph G have a minimum distance spanning tree T that is also distance preserving?*

Question 2. *If the graph G has a minimum distance spanning tree T that is distance preserving, is it distance preserving from a vertex v for which $\sigma_G(v)$ is minimum?*

6 $\mathbf{Q_n}$

Given a nontrivial connected graph G define $s(G) := \min_T \sigma(T)/\sigma(G)$ where T is a spanning tree of G. In view of Theorem K we know that $1 \leq s(G) < 2$ for every connected graph. If it is known a priori that $s(G)$ is nearly 2 for some graph G there may be no need to find a spanning tree with least distance since any spanning tree found using Algorithms 1 or 2 will have approximately the same distance. This notion underlies the following.

Project. *Find graphs G for which $s(G) > 2(1 - \varepsilon)$.*

Since $K_{1,n-1}$ is the tree of order n with minimum distance,

$$s(K_n) = (n-1)^2 / \binom{n}{2} = 2(1 - 1/n).$$

Consequently $s(K_n) \sim 2$. On the other hand, we have

$$s(C_n) = \frac{\sigma(P_n)}{\sigma(C_n)} = \binom{n+1}{3} / \frac{n}{2} \left\lfloor \frac{n^2}{4} \right\rfloor = \frac{4}{3}$$

for odd n.

For our final example we first need the following.

Theorem L. (Burns and Entringer [3]) *Let $G := K_{n_1,n_2,\ldots,n_r}$ be a complete r-partite graph of order n with r-partition $\{V_1, V_2, \ldots, V_r\}$, $r \geq 2$, satisfying $|V_i| = n_i$, $1 \leq i \leq r$, and $n_1 \leq n_2 \leq \cdots \leq n_r$. A spanning tree T of G is a minimum distance spanning tree iff $n_1 = 1$ and $T = K_{1,n-1}$ or $n_1 > 1$ and T is the spanning tree with exactly two internal vertices, one with degree $n - n_1$ and the other with degree n_1. This tree has distance $n^2 - 3n + 2 + nn_1 - n_1^2$.*

Letting T be the tree defined in Theorem L for the case $r = 2$ and $n_1 = n_2 = m$ gives

$$s(K_{m,m}) = \frac{\sigma(T)}{\sigma(K_{m,m})} = \frac{5m^2 - 6m + 2}{3m^2 - 2m} \sim \frac{5}{3}.$$

It is not clear from these three examples how the value of $s(G)$ depends on the density of G. Although we do not know of any class of sparse graphs G_n satisfying $s(Gn) \sim 2$, we will present speculation that the hypercubes, Q_n, form such a class. So let us focus on the following.

Problem. *Find a minimum distance spanning tree of Q_n.*

Recognizing that $Q_n = K_2 \times Q_{n-1}$ we first consider the more general problem of constructing a minimum distance spanning tree T^* of $G = K_2 \times H$ from a minimum distance spanning tree T of the connected graph H of order n. Two techniques suggest themselves.

Type I: (See Figure 8) Suppose $G = K_2 \times H$ and T is a minimum distance spanning tree of H. To form a Type I spanning tree T^* of $K_2 \times H$ simply append one end vertex to each vertex of T. Calling this set of appended vertices S we have

$$\begin{aligned}\sigma(T^*) &= \sum_{\{u,v\}\subseteq V(T)} d(u,v) + \sum_{\substack{u\in V(T)\\ v\in S}} d(u,v) + \sum_{\{u,v\}\subseteq S} d(u,v)\\ &= \sigma(T) + \sum_{\substack{u\in V(T)\\ v\in V(T)}} [d(u,v)+1] + \sum_{\{u,v\}\subseteq V(T)} [d(u,v)+2]\\ &= 4\sigma(T) + 2n^2 - n.\end{aligned}$$

Type II: (See Figure 9) Suppose $G = K_2 \times H$ and T is a minimum distance spanning tree of H. To form a Type II spanning tree T^* of $K_2 \times H$ take two disjoint copies of T, say T and T', choose centroid vertices v and v' in

T and T', respectively, and join these vertices with an edge. We have

$$\begin{aligned}\sigma(T^*) &= \sum_{\{u,w\}\subseteq V(T)} d(u,w) + \sum_{\substack{u\in V(T)\\ w\in V(T')}} d(u,w) + \sum_{\{u,w\}\subseteq V(T')} d(u,w)\\ &= 2\sigma(T) + \sum_{\substack{u\in V(T)\\ w\in V(T')}} [d(u,v)+1+d(v',w]\\ &= 2\sigma(T) + n\sigma_T(v) + n^2 + n\sigma_T(v') = 2\sigma(T) + 2n\sigma_T(v) + n^2.\end{aligned}$$

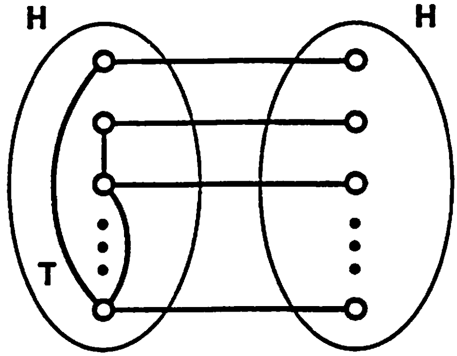

Figure 8. Type I spanning tree T of $H \times K_2$

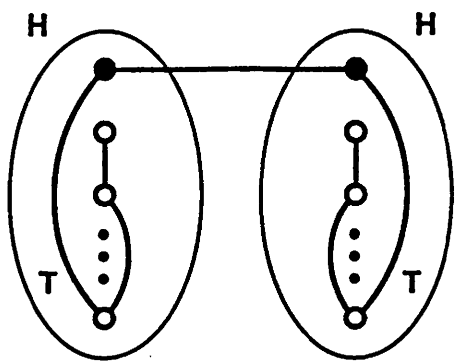

Figure 9. Type II spanning tree T^* of $H \times K_2$

As the following examples show, neither the Type I tree nor the Type II tree necessarily has smaller distance than the other.

Letting T_{I}^* and T_{II}^* be Type I and Type II, respectively, spanning trees of $K_2 \times P_n$ we obtain

$$\sigma(T_{\text{I}}^*) = 4\binom{n+1}{3} + 2n^2 - n = \frac{2n^3+6n^2-5n}{3}$$

while

$$\sigma(T_{\mathrm{II}}^*) = 2\binom{n+1}{3} + 2n\left\lfloor \frac{n^2}{4} \right\rfloor + n^2 \geq \frac{5n^3+6n^2-5n}{6} > \sigma(T_{\mathrm{I}}^*)$$

for $n \geq 6$.

On the other hand, letting T_{I}^* and T_{II}^* be Type I and Type II, respectively, spanning trees of $K_2 \times K_{1,n-1}$ we obtain

$$\sigma(T_{\mathrm{I}}^*) = 4(n-1)^2 + 2n^2 - n = 6n^2 - 9n + 4$$

while

$$\sigma(T_{\mathrm{II}}^*) = 2(n-1)^2 + 2n(n-1) + n^2 = 5n^2 - 6n + 2 < \sigma(T_{\mathrm{I}}^*)$$

for $n \geq 3$.

Equality is also possible. Let $H = Q_n$ (so that H has order 2^n). Solving the recurrence relation for the distance of the Type I spanning tree T_n of Q_n results in

$$\sigma(T_n) = \frac{n-1}{2}4^n + 2^{n-1}.$$

Now let T_n be a Type II spanning tree of Q_n and denote a centroid vertex of it by v_n. Then $\sigma(v_n) = n2^{n-1}$ since it satisfies the recurrence relation $\sigma(v_{n+1}) = 2\sigma(v_n) + 2^n$. Next, the recurrence relation $\sigma(T_{n+1}) = 2\sigma(T_n) + (n+1)4^n$ gives

$$\sigma(T_n) = \frac{n-1}{2}4^n + 2^{n-1}.$$

Thus the Type I and Type II trees have the same distance when $G = Q_n$. In fact, more is true.

Theorem M. (McCanna [18]) *The Type I and Type II spanning trees are the same for Q_n.*

Empirical results for small n suggest the following.

Conjecture. *The tree defined above is a minimum distance spanning tree of Q_n, $n \geq 1$.*

If this conjecture is true, then we would have

$$s(Q_n) = 2\left(1 - \frac{1}{n}\right) + \frac{1}{n2^{n-1}} \sim 2.$$

7 Other results

Plesnik [19] conjectured that every rational $r \geq 1$ is the mean distance of some graph; this was proved by Truszcynski [22] and independently by

Hendry [14]. Later, Hendry [15] showed that given any $r \geq 2$ and $\varepsilon > 0$ there is a tree T with $\lfloor \mu(T) - r \rfloor < \varepsilon$. He then asked whether for rational $r \geq 2$ there is always a tree whose average distance is exactly r and Winkler responded with the following.

Theorem N. (Winkler [29]) *Let a rational number r be given. If $r > 2$, then there are a (countably) infinite number of nonisomorphic trees whose mean distance is exactly r. If $r < 2$ then there is a tree whose mean distance is r just when r can be written in the form $2 - 2/k$, where k is an integer greater than 1; if $k = \binom{m}{2}$ for some integer $m > 3$ or if $k = 30$, there are exactly two such trees, otherwise only one. Finally, there are two trees whose mean distance is 2.*

Gutman, Yeh, and Chen also have studied the distances realizable by graphs. In particular, they offer the following.

Conjecture. (Gutman,Yeh, and Chen [13]) *There are only finitely many positive integers that are not the distance of some tree.*

Winkler proposed the following (all graphs in the remainder of this section are assumed to be nontrivial).

Conjecture. (Winkler [28]) *Every connected graph G has a vertex v for which $\mu(G - v)/\mu(G) < 4/3$.*

A weakened form of this conjecture has been proved.

Theorem O. (Bienstock and Győri [2]) *Every connected graph G has a vertex v for which $\mu(G - v)/\mu(G) < 4/3 + O(n^{-1/5})$.*

A strengthened form of Winkler's edge analog of his latter conjecture was also obtained.

Theorem P. (Bienstock and Győri [2]) *Every connected graph G has an edge e for which $\frac{\sigma(G-e)}{\sigma(G)} \leq \frac{4}{3}$.*

The result $s(C_n) = 4/3$ from section 6 shows that the theorem is sharp for odd n.

8 Concluding remarks

There are variants in terminology used by others. For examples, distance of a graph has been called Wiener index and centroid vertices have been called median vertices. In general median vertices of a connected graph are defined to be the vertices with minimum distance.

The distance of graphs similar to those in the class $\mathcal{T}(n, \Delta)$ of section 4 have been studied by Gutman, Yeh, and Chen [11].

As pointed out by Wiener [23], if T is a tree then $\sigma(T)$ can also be calculated by first calculating a weight $p(e)$ for each edge e and then summing these weights over all edges. The weight of each edge $e = uv$ is just the product of the number of vertices closer to u and the number of vertices closer to v, i.e., it is the product of the orders of the two components of $T - uv$. It is easy to see that the resultant sum is $\sigma(T)$. This has been noticed by Bienstock and Gyóri [2], Gutman [10], and others. The weight $p(e)$ has been called the *path number of the edge* e and an analogous definition has been given for the *path number of a vertex* v, i.e., the number of paths in the tree containing v as an internal vertex. This latter concept has been explored at some length by Burns and the author [3].

We have limited the bibliography mainly to articles dealing with distance in trees. To list all papers concerned with distance in graphs would be impractical. The reader wishing to study distance in the wider context should consult the bibliographies of references [12] and [19] and the conjectures of the *Graffiti* program of Fajtlowicz. Those wishing to study the many applications of graph distance in chemistry can find further materials in references [4], [12], and [20].

References

[1] CA. Barefoot, R.C. Entringer, and L.A. Szèkely, Extremal ratios for distance in trees, (in preparation).

[2] D. Bienstock and E. Gyóri, Average distance in graphs with removed elements, *J. Graph Theory* **12** (1988), 375–390.

[3] K. Burns and R.C. Entringer, A graph-theoretic view of the United States Postal Service, in *Graph Theory, Combinatorics, and Algorithms: Proceedings of the Seventh Quadrennial International Conference on the Theory and Applications of Graphs* (Y. Alavi and A. Schwenk, eds.), John Wiley, New York (1995), 323–334.

[4] E.R. Canfield, R.W. Robinson, and D.H. Rouvray, Determination of the Wiener molecular branching index for the general tree, *J. Comput. Chem.* **6** (1985), 598–609.

[5] R.C. Entringer, D.E. Jackson, and D.A. Snyder, Distance in graphs, *Czechoslovak Math. J.* **26**(101) (1976), 283–296.

[6] R.C. Entringer, A. Meir, J.W. Moon, and L.A. Szèkely, On the Wiener index of trees from certain families, *Australas. J. Combin.* **10** (1994), 211–224.

[7] R.C. Entringer, D.J. Kleitman, and L.A. Szèkely, A note on spanning trees with minimum average distance, *Bull. Inst. Combin. Appl.* **17** (1996), 71–78.

[8] R.C. Entringer, Bounds for the average distance – inverse degree product for trees, (in preparation).

[9] O. Gerstel and S. Zaks, A new characterization of tree medians with applications to distributed algorithms, *Graph-theoretic concepts in computer science* (Weisbaden-Naurod, 1992), 135–144, *Lecture Notes in Comput. Sci.*, **657**, Springer, Berlin, 1993.

[10] I. Gutman, A new method for the calculation of the Wiener number of acyclic molecular graphs, preprint.

[11] I. Gutman, Y.N. Yeh, S.L. Lee, and J.C. Chen, Wiener number of dendrimers, preprint.

[12] I. Gutman, Y.N. Yeh, S.L. Lee, and Y.L. Lou, Recent results in the theory of the Wiener number, preprint.

[13] I. Gutman, Y.N. Yeh, and J.C. Chen, On the sum of all distances in graphs, *Tamkang J. Math.* **25** (1994), 83–86.

[14] G.R.T. Hendry, Existence of graphs with prescribed mean distance, *J. Graph Theory* **10** (1986), 173–175.

[15] G.R.T. Hendry, On mean distance in certain classes of graphs, *Networks* **19** (1989), 451–457.

[16] D.S. Johnson, J.K. Lenstra, and A.H.G. Rinnooy-Kan, The complexity of the network design problem, *Networks* **8** (1978), 279–285.

[17] C. Jordan, Sur les assemblages de lignes, *J. Reine Angew. Math.* **70** (1869), 185–190.

[18] J. McCanna, unpublished.

[19] J. Plesnik, On the sum of all distances in a graph or digraph, *J. Graph Theory* **8** (1984), 1–21.

[20] D.H. Rouvray, Prediction chemistry from topology, *Scientific American* **25** (1986), 40–47.

[21] R. Shi, The average distance of trees, *Systems Sci. Math. Sci.* **6** (1993), 18–24.

[22] M. Truszcynski, A graph with mean distance being a given rational, *Demonstratio Math.* **18** (1985), 619–620.

[23] H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947), 17–20 .

[24] H. Wiener, Characterization of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons, *J. Am. Chem. Soc.* **69** (1947), 2636–2638.

[25] H. Wiener, Influence of interatomic forces on paraffin properties, *J. Chem. Phys.* **15** (1947), 766.

[26] H. Wiener, Vapor pressure-temperature relationships among the branched paraffin hydrocarbons, *J. Phys. Chem.* **52** (1948), 425–430.

[27] H. Wiener, Relation of the physical properties of the isometric alkanes to molecular structure, *J. Phys. Chem.* **52** (1948), 1082–1089.

[28] P. Winkler, Mean distance and the 'four-thirds conjecture', *Congr. Numer.* **54** (1986), 63–72.

[29] P. Winkler, Mean distance in a tree, *Discrete Appl. Math.* **27** (1990), 179–185.

[30] B. Zelinka, Medians and peripherians of trees, *Arch. Math.* (Brno) **4** (1968), 87–95.