involve
a journal of mathematics

# Optimal trees for functions of internal distance

Alex Collins, Fedelis Mutiso and Hua Wang

msp

2012 vol. 5, no. 3

INVOLVE 5:3 (2012)
dx.doi.org/10.2140/involve.2012.5.371

# Optimal trees for functions of internal distance

Alex Collins, Fedelis Mutiso and Hua Wang

(Communicated by Jerrold Griggs)

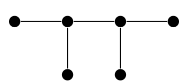
The sum of distances between vertices of a tree has been considered from many aspects. The question of characterizing the extremal trees that maximize or minimize various such “distance-based” graph invariants has been extensively studied. Such invariants include, to name a few, the sum of distances between all pairs of vertices and the sum of distances between all pairs of leaves. With respect to the distances between internal vertices, we provide analogous results that characterize the extremal trees that minimize the value of any nonnegative and nondecreasing function of internal distances among trees with various constraints.

## 1. Introduction

As a classic example of the distance-based graph invariants, the *Wiener index* [1947] is one of the most well used chemical indices that correlate a chemical compound’s structure (the “molecular graph”) with experimentally gathered data of the compound’s physical-chemical properties such as boiling point, surface pressure, etc. The Wiener index is defined as

$$W(G) = \sum_{\{u,v\}\subseteq V(G)} d(u, v),$$

where $d(u, v)$ is the distance between two vertices $u$ and $v$ and the sum is over all pairs of vertices. For example, the tree shown here has index 29:



The extremal trees that maximize or minimize the Wiener index among general trees [Dobrynin et al. 2001], trees with a given maximum degree [Fischermann et al. 2002], and trees with given degree sequence [Zhang et al. 2008; 2010] have been characterized through various approaches. A general approach was presented dealing with functions of distances between vertices [Schmuck et al. 2012].

*MSC2010:* primary 05C05, 05C12; secondary 05C30.
*Keywords:* internal distances, trees, extremal.
This work was partially supported by a grant from the Simons Foundation (#245307 to Hua Wang).

Recently, the *gamma index* [Székely et al. 2011], also known as the *terminal Wiener index* [Gutman et al. 2009], was introduced due to its applications in phylogenetic reconstruction and biochemistry. For a tree $T$, the gamma index is defined as the sum of distances between all pairs of leaves. It is interesting to note that the star minimizes both the Wiener index and the gamma index among trees of given order. Among trees of a given degree sequence, the "greedy tree" (Definition 1.1) was shown to minimize both the Wiener index [Zhang et al. 2008] and the gamma index [Székely et al. 2011].

**Definition 1.1** (greedy trees). With given vertex degrees, the *greedy tree* is achieved through the following *greedy algorithm*:

(i) Label the vertex with the largest degree as $v$ (the root).

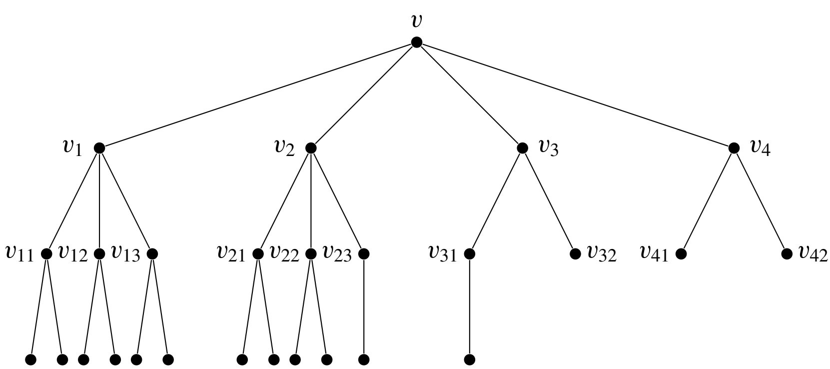
(ii) Label the neighbors of $v$ as $v_1, v_2, \ldots$, and assign the largest degrees available to them such that $\deg(v_1) \geq \deg(v_2) \geq \cdots$.

(iii) Label the neighbors of $v_1$ (except $v$) as $v_{11}, v_{12}, \ldots$ such that they take all the largest degrees available and that $\deg(v_{11}) \geq \deg(v_{12}) \geq \cdots$, then do the same for $v_2, v_3, \ldots$.

(iv) Repeat (iii) for all the newly labeled vertices, always starting with the neighbors of the labeled vertex with largest degree whose neighbors are not labeled yet.

For example, here is a greedy tree with degree sequence

$$\{4, 4, 4, 3, 3, 3, 3, 3, 3, 3, 2, 2, 1, \ldots, 1\}.$$




**Theorem 1.2** [Schmuck et al. 2012]. *Let $f(x)$ be any nonnegative, nondecreasing function of $x$. Then the graph invariant*

$$W_f(T) = \sum_{\{u,v\} \subseteq V(T)} f(d(u,v))$$

*is minimized by the greedy tree among all trees with given degree sequence.*

Theorem 1.2 immediately implies the extremality of the greedy tree regarding many different distance-based graph invariants. Take, for instance, the Wiener index ($f(x)=x$), the hyper-Wiener index ($f(x)=x(x+1)/2$), and the generalized Wiener index ($f(x)=x^\alpha$). See [Schmuck et al. 2012] for more details.

Following the Wiener index and the gamma index, a natural next step is to consider the sum of distances between internal vertices. In [Székely and Wang 2005], it was asked if the extremal values of the sums of distances between internal vertices, between leaves, or between internal vertices and leaves can be explored through a similar approach. The sum of distances between internal vertices was brought up again in [Bartlett et al. 2013] and named the *spinal index*:

$$S(T)=\sum_{\{u,v\}\subseteq V(T)-L(T)} d(u,v),$$

where $L(T)$ denotes the set of leaves of $T$. The extremal trees that maximize or minimize the spinal index have been studied based on the known results regarding the Wiener index [Bartlett et al. 2013]. Similar to $W_f(T)$, it is natural to consider the *spinal function index*, defined as

$$S_f(T)=\sum_{u,v\in V(T)-L(T)} f(d(u,v))$$

for any nonnegative, nondecreasing function $f$.

The goal of this note is to show that one can provide general statements analogous to Theorem 1.2 and its consequences for $S_f(T)$. By establishing Proposition 2.4, we characterize the trees that minimize the spinal function index among trees with given order and number of leaves (Theorem 3.2), with given degree sequence (Theorem 3.4), as well as with given order and maximum degree (Theorem 3.5).

## 2. Preliminaries

Our study consists of a combination of techniques in [Bartlett et al. 2013] and [Schmuck et al. 2012]. We first recall the following crucial result, where $p_k(T)$ is the number of pairs $(u,v)$ of vertices such that $d(u,v)\le k$.

**Theorem 2.1** [Schmuck et al. 2012]. *Let $d_1\ge d_2\ge\cdots\ge d_n$ be positive integers such that $\sum_i d_i=2(n-1)$, and let $k$ be another arbitrary positive integer. Among all trees with degree sequence $(d_1,d_2,\dots,d_n)$, the greedy tree maximizes $p_k(T)$.*

**Remark 2.2.** Theorem 2.1 implies Theorem 1.2. Indeed, note that

$$W_f(T)=\sum_{k\ge 0}\bigl(f(k+1)-f(k)\bigr)\Bigl|\Bigl\{\{u,v\}\subseteq V(T)\;\Big|\; d(u,v)>k\Bigr\}\Bigr|,$$

and that $f(k)-f(k-1)$ is nonnegative for all $k$ (we set $f(0)=0$).

The idea of comparing greedy trees with different degree sequences through "majorization" was used in [Zhang et al. 2012], where the following is defined.

Consider two nonincreasing sequences $\pi = (d_0, \dots, d_{n-1})$, $\pi' = (d'_0, \dots, d'_{n-1})$. If

$$\sum_{i=0}^{k} d_i \le \sum_{i=0}^{k} d'_i \quad \text{for } k = 0, \dots, n-2 \quad \text{and} \quad \sum_{i=0}^{n-1} d_i = \sum_{i=0}^{n-1} d'_i,$$

then $\pi'$ is said to *majorize* the sequence $\pi$, denoted by

$$\pi \triangleleft \pi'.$$

**Lemma 2.3** [Wei 1982]. *Let* $\pi = (d_0, \dots, d_{n-1})$ *and* $\pi' = (d'_0, \dots, d'_{n-1})$ *be two nonincreasing graphic degree sequences. If* $\pi \triangleleft \pi'$, *then there exists a series of graphic degree sequences* $\pi_1, \dots, \pi_m$ *such that* $\pi \triangleleft \pi_1 \triangleleft \cdots \triangleleft \pi_m \triangleleft \pi'$, *where* $\pi_i$ *and* $\pi_{i+1}$ *differ at exactly two entries, say* $d_j$ $(d'_j)$ *and* $d_l$ $(d'_l)$ *of* $\pi_i$ $(\pi_{i+1})$, *with* $d'_j = d_j + 1$, $d'_l = d_l - 1$ *and* $j < l$.

With Lemma 2.3, the following can be shown in a way similar to Theorem 2.4 in [Zhang et al. 2012].

**Proposition 2.4.** *For two different degree sequences* $\pi$ *and* $\pi'$, *if* $\pi \triangleleft \pi'$, *then*

$$p_k(T^*_\pi) \le p_k(T^*_{\pi'})$$

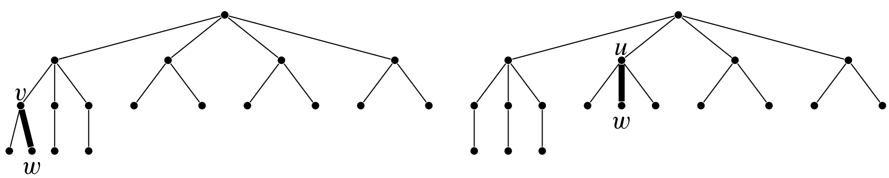
*for any* $k \ge 1$ *where* $T^*_\pi$ *and* $T^*_{\pi'}$ *are the greedy trees with degree sequences* $\pi$ *and* $\pi'$, *respectively.*

*Proof.* By Lemma 2.3, it is sufficient to show the statement for degree sequences

$$\pi = (d_0, \dots d_{n-1}) \triangleleft (d'_0, \dots, d'_{n-1}) = \pi'$$

that differ only at the $j$-th and $l$-th entries with $d'_j = d_j + 1$, $d'_l = d_l - 1$ for some $j < l$.

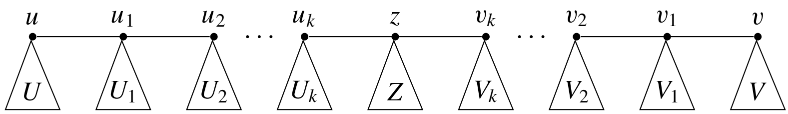
Let $T'_\pi$ be the tree constructed from $T^*_\pi$ by removing the edge $vw$ and adding an edge $uw$, where $u$ and $v$ are the vertices corresponding to $d_j$ and $d_l$, respectively, and $w$ is a child of $v$:




$T^*_\pi$, $\pi = (4, 4, 3, 3, 3, 3, 2, 2, 1, \dots, 1)$ $\quad$ $T'_\pi$, $\pi' = (4, 4, 4, 3, 3, 2, 2, 2, 1, \dots, 1)$

Let $T'$ be the tree obtained from $T^*_\pi$ after removing $w$ and its descendants. Then the next claim follows from the structure of the greedy tree $T^*_\pi$ (see, for instance, [Wang 2008; Zhang et al. 2008; 2012]).

**Claim 2.5.** *Let the path from $u$ to $v$ be $uu_1u_2\cdots u_m(z)v_m\cdots v_2v_1v$, where the existence of $z$ depends on the parity of $d(u,v)$. Let $U$, $U_i$, $V$, $V_i$, $Z$ denote the component containing $u$, $u_i$, $v$, $v_i$, $z$, respectively, after removing the edges on this path from $T'$:*




*Then $p_k(U,u)\geq p_k(V,v)$ and $p_k(U_i,u_i)\geq p_k(V_i,v_i)$ for any $1\leq i\leq m$ and any $k\geq 1$. Here $p_k(T,x)$ denotes the number of vertices $y\in V(T)$ such that $d(x,y)\leq k$.*

In particular, Claim 2.5 implies that, for any $k\geq 1$, there are more vertices within distance $k$ from $u$ in $T'$ than those from $v$.

Now simple calculations (see [Wang 2008], for instance) show that

$$p_k(T^*_{\pi'})\geq p_k(T'_\pi)\geq p_k(T^*_\pi)\quad\text{for any }\ k\geq 1.\qquad\square$$

## 3. Extremal trees with respect to $S_f(T)$

First note that any tree $T$ that is not a star has at least two internal vertices. Hence $S_f(T)\geq 0$ for any $T$. The following observation is trivial.

**Proposition 3.1.** *Among all trees with the same order, the star has the minimal $S_f(T)=0$.*

As shown in Remark 2.2, we only need to focus on $p_k(T)$ for other more involved cases. In what follows we show that several statements from [Bartlett et al. 2013] can be easily generalized for $S_f(T)$. It is worth pointing out that, with the understanding of the preliminaries (particularly with Proposition 2.4), these results can be obtained in a very similar fashion as [Bartlett et al. 2013].

**Theorem 3.2.** *For a tree $T$ with given order and number of leaves, let $T^*_\pi$ denote a greedy tree with degree sequence*

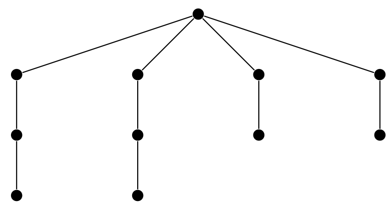
$$\pi=\big(|L(T)|,2,\ldots,2,\underbrace{1,\ldots,1}_{|L(T)|\text{ 1's}}\big).$$

*Then $p_k(T^*_\pi)\geq p_k(T)$ for any $k\geq 1$. Hence $S_f(T)$ is minimized by the same tree by Remark 2.2.*

**Remark 3.3.** Such a tree is called "star-like", achieved by attaching exactly one pendant edge to each of the leaves of a greedy tree with degree sequence

$$(|L(T)|,2,\ldots,2,1,\ldots,1).$$

Here is an example:



*Proof.* Consider the subtree $T'$ induced by the internal vertices of $T$. We have that $p_k(T')$ is maximized by a greedy tree with $|V(T)| - |L(T)|$ vertices and at most $|L(T)|$ leaves (each of the leaves in $T'$ has at least one vertex in $L(T)$ as a neighbor in $T$).

Among the degree sequences of such trees,

$$\Big(|L(T)|, \underbrace{2, \dots, 2,}_{|V(T)|-2|L(T)|-1 \text{ 2's}} \underbrace{1, \dots, 1}_{|L(T)| \text{ 1's}}\Big)$$

majorizes all others. Then $T$ is a greedy tree with degree sequence

$$\Big(|L(T)|, \underbrace{2, \dots, 2,}_{|V(T)|-|L(T)|-1 \text{ 2's}} \underbrace{1, \dots, 1}_{|L(T)| \text{ 1's}}\Big). \qquad \square$$

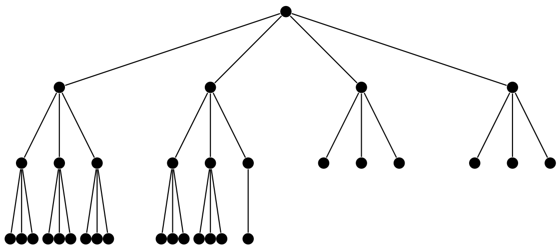
**Theorem 3.4.** *Among trees with given order and degree sequence, $p_k(T)$ is maximized by the greedy tree. Consequently, $S_f(T)$ is minimized by the greedy tree.*

*Proof.* First note that with given degree sequence, $|L(T)|$ is determined.

To minimize $S_f(T)$, note that $p_k(T')$ is minimized by a greedy tree with the degree sequence of $T'$. Let the degree sequence of $T$ be $(d_1, d_2, \dots)$. Then the degree sequence of $T'$ is $(d_1 - k_1, d_2 - k_2, \dots)$ where $k_i \geq 0$ is the number of leaf-neighbors of the vertex corresponding to the degree $d_i$. The degree sequence (of $T'$) of this form that majorizes all others is when $k_1 = k_2 = \cdots = k_i = 0$ for $i$ as large as possible. Note that this is the case only when all the vertices (in $T$) of large degrees have no leaf-neighbors, or in other words, the leaves of $T$ are adjacent only to (as few as possible) internal vertices of the smallest degrees in $T$. This happens only if $T$ is the greedy tree. Thus the conclusion follows from Proposition 2.4. $\square$

The *complete $k$-ary tree* with a given maximum degree $k$ is defined in a similar way as the greedy tree, except that the vertices $v, v_1, \dots$ take the maximum degree $k$ until there are not enough vertices (see figure on the next page). As a result, the complete $k$-ary tree has degree sequence $(k, k, \dots, k, m, 1, \dots, 1)$ for some $1 < m \leq k$.

The extremality of the complete $k$-ary tree follows in the same way as previous arguments.



A complete 4-ary tree.

**Theorem 3.5.** *Among trees with given order and maximum degree $k$, $p_k(T)$ is maximized and $S_f(T)$ is minimized by the complete $k$-ary tree.*

## 4. Concluding remarks

We have shown, for any nonnegative, nondecreasing function $f$, that the sum of $f(d(u, v))$ over all pairs of internal vertices is minimized by the same trees as the ones that minimize the original spinal index. The analogue can be easily obtained for nonincreasing functions. These results, providing a much stronger generalization on this study, were obtained by utilizing tools from previous studies. We also hope that we have illustrated the power of the established approaches in the study of such extremal graphs with respect to distance-based graph invariants.

## Acknowledgments

The authors thank the referee for many helpful suggestions that improved the presentation of this note.

## References

[Bartlett et al. 2013] M. Bartlett, E. Krop, C. Magnant, F. Mutiso, and H. Wang, "Variations of distance-based invariants of trees", *J. Combin. Math. Combin. Comput.* (2013). To appear.

[Dobrynin et al. 2001] A. A. Dobrynin, R. Entringer, and I. Gutman, "Wiener index of trees: theory and applications", *Acta Appl. Math.* **66**:3 (2001), 211–249. MR 2002i:05035 Zbl 0982.05044

[Fischermann et al. 2002] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, and L. Volkmann, "Wiener index versus maximum degree in trees", *Discrete Appl. Math.* **122**:1-3 (2002), 127–137. MR 2003d:05061 Zbl 0993.05061

[Gutman et al. 2009] I. Gutman, B. Furtula, and M. Petrović, "Terminal Wiener index", *J. Math. Chem.* **46**:2 (2009), 522–531. MR 2011e:05075 Zbl 05601386

[Schmuck et al. 2012] N. S. Schmuck, S. G. Wagner, and H. Wang, "Greedy trees, caterpillars, and Wiener-type graph invariants", *MATCH Commun. Math. Comput. Chem.* **68**:1 (2012), 273–292. MR 2986487

[Székely and Wang 2005] L. A. Székely and H. Wang, "On subtrees of trees", *Adv. in Appl. Math.* **34**:1 (2005), 138–155. MR 2005h:05050 Zbl 1153.05019

[Székely et al. 2011] L. A. Székely, H. Wang, and T. Wu, "The sum of the distances between the leaves of a tree and the 'semi-regular' property", *Discrete Math.* **311**:13 (2011), 1197–1203. MR 2012d:05106 Zbl 1222.05027

[Wang 2008] H. Wang, "The extremal values of the Wiener index of a tree with given degree sequence", *Discrete Appl. Math.* **156**:14 (2008), 2647–2654. MR 2009i:05076 Zbl 1155.05020

[Wei 1982] W. D. Wei, "The class $\mathfrak{A}(R, S)$ of (0, 1)-matrices", *Discrete Math.* **39**:3 (1982), 301–305. MR 84j:05029 Zbl 0484.15015

[Wiener 1947] H. Wiener, "Structural determination of paraffin boiling points", *Journal of the American Chemical Society* **69**:1 (1947), 17–20.

[Zhang et al. 2008] X.-D. Zhang, Q.-Y. Xiang, L.-Q. Xu, and R.-Y. Pan, "The Wiener index of trees with given degree sequences", *MATCH Commun. Math. Comput. Chem.* **60**:2 (2008), 623–644. MR 2009i:05078 Zbl 1195.05022

[Zhang et al. 2010] X.-D. Zhang, Y. Liu, and M.-X. Han, "Maximum Wiener index of trees with given degree sequence", *MATCH Commun. Math. Comput. Chem.* **64**:3 (2010), 661–682. MR 2012d:05130 Zbl 06124037

[Zhang et al. 2012] X.-M. Zhang, X.-D. Zhang, D. Gray, and H. Wang, "The number of subtrees of trees with given degree sequence", *J. Graph Theory* (2012).

Received: 2012-11-05 Revised: 2013-03-10 Accepted: 2013-03-30

acollins38@gsu.edu *Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, United States*

fm00344@georgiasouthern.edu *Department of Mathematical Sciences, Georgia Southern University, Statesboro, GA 30460, United States*

hwang@georgiasouthern.edu *Department of Mathematical Sciences, Georgia Southern University, Statesboro, GA 30460, United States*

# involve

msp.org/involve





See inside back cover or msp.org/involve for submission instructions. The subscription price for 2012 is US $105/year for the electronic version, and $145/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

involve

2012 vol. 5 no. 3