VARIATIONS OF DISTANCE-BASED INVARIANTS OF TREES

MARCUS BARTLETT, ELLIOT KROP, COLTON MAGNANT, FEDELIS MUTISO, AND HUA WANG

ABSTRACT. Introduced in 1947, the Wiener index (sum of distances between all pairs of vertices) is one of the most studied chemical indices. Extensive results regarding the extremal structure of the Wiener index exist in the literature. More recently, the Gamma index (also called the Terminal Wiener index) was introduced as the sum of all distances between pairs of leaves. It is known that these two indices coincide in their extremal structures and that a nice functional relation exists for k-ary trees but not in general. In this note, we consider two natural extensions of these concepts, namely the sum of all distances between internal vertices (the Spinal index) and the sum of all distances between internal vertices and leaves (the Bartlett index). We first provide a characterization of the extremal trees of the Spinal index under various constraints. Then, its relation with the Wiener index and Gamma index is studied. The functional relation for k-ary trees also implies a similar result on the Bartlett index.

1. INTRODUCTION

Chemical indices have been introduced by chemists to 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* [12] of a graph G is one of the most well studied such indices, 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 unordered pairs of vertices. See Figure 1 for an example.

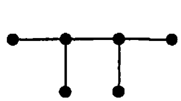


FIGURE 1. A tree with Wiener index 29

JCMCC 91 (2014), pp. 19-29

Ever since its introduction in 1947, the Wiener index has been extensively studied by both chemists and mathematicians. In particular, the extremal trees that maximize or minimize the Wiener index among general trees [2], trees with a given maximum degree [3], and trees with given degree sequence [14] have been characterized through various approaches. Most recently, a general approach was presented dealing with functions of distances between vertices [6].

The *Gamma index* [8], also known as the *terminal Wiener index* [5], was introduced recently due to its applications in phylogenetic reconstruction and biochemistry. For a tree T, the Gamma index is defined as

$$\Gamma(T) = \sum_{\{u,v\} \subseteq L(T)} d(u,v)$$

where $L(T)$ is the set of leaves of T. It is not difficult to notice the similarity between $\Gamma(T)$ and $W(T)$ for a tree. Indeed, the star minimizes both indices among trees of given order. Among trees of a given degree sequence, the "greedy tree" (Definition 1) was shown to minimize both the Wiener index [14] and the Gamma index [8]. In [9], a simple example shows that there is no functional relation between these two indices in general. However, also in [9], for two k-ary trees T and T', the following is shown

$$W(T) - W(T') = \left(\frac{k-1}{k-2}\right)^2 (\Gamma(T) - \Gamma(T')). \qquad (1)$$

As other variations of distance-based graph invariants such as $W(T)$ or $\Gamma(T)$, it is natural to consider the sum of the distances between internal vertices and the distances between internal vertices and leaves. We define the former as the *Spinal index*

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

and the latter as the *Bartlett index*

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

It is easy to notice that, for any tree T,

$$W(T) = \Gamma(T) + S(T) + B(T). \qquad (2)$$

It is also clear that all these indices tend to be larger when T is sparse (i.e. vertices are far from each other) and smaller when T is compact (i.e. vertices are close to each other).

In this note, we first consider the extremal structures with respect to $S(T)$ in Section 2. As one may expect, these structures coincide with what is known for $W(T)$. Such coincidence motivates the study of the relation between the classical Wiener index and $S(T)$. In Section 3, simple examples show that there is no general functional relation between either $S(T)$ or $B(T)$ and $W(T)$. However, for k-ary trees, we can obtain results analogous to (1) for $S(T)$ and then for $B(T)$ as a corollary. In Section 4, we provide a brief summary and raise some questions.

2. Some extremal results

The star and path are extremal among general trees of given order with respect to numerous graph invariants, among which distance-based indices. In particular, $W(T)$ is maximized by a path and minimized by a star among trees of given order.

Now consider the distances between internal vertices. It is important to note that

$$S(T) = W(T') \tag{3}$$

where T' is the subtree of T induced by $V(T) - L(T)$.

Any tree T that is not a star has at least two internal vertices and hence $S(T) > 0$. Since $S(T) = 0$ for a star, we have the following observation.

Proposition 2.1. *Among trees with given order, the star is the unique tree that minimizes* $S(T)$.

Similarly, a tree T has at most $|V(T)| - 2$ internal vertices (with the upper bound achieved if and only if T is a path) and $S(T) = W(T')$ is maximized when T' is a path.

Proposition 2.2. *Among trees with given order, the path is the unique tree that maximizes* $S(T)$.

It is easy to see that the star has the largest number of leaves and largest possible maximum degree while the path has the smallest possible number of leaves and the smallest possible degrees. Then it is natural to consider the extremal questions with restrictions on $|L(T)|$ or the degrees. Indeed such questions have been explored for many other indices including the Wiener index. It is known [8, 14] that both $W(T)$ and $\Gamma(T)$ are minimized by the so called "greedy tree" (Definition 1) among trees of a given degree sequence. We list the definition here for completeness.

Definition 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, ..., 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}, ... 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, ...;*

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

For example, Figure 2 displays a greedy tree with degree sequence $(4,4,4,3,3,3,3,3,3,3,2,2,1,\ldots,1)$.

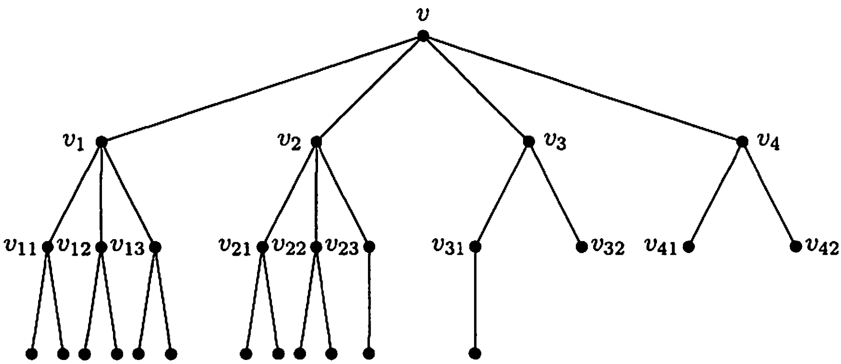


FIGURE 2. A greedy tree.

It immediately follows from (3) that $S(T)$ must be minimized by a tree T if T' is a greedy tree. Before getting to further conclusions, we first introduce the concept of "majorization" and a technical result similar to one in [15].

Consider two nonincreasing sequences $\pi = (d_0, \cdots, d_{n-1})$ and $\pi' = (d'_0, \cdots, d'_{n-1})$, if

$$\sum_{i=0}^{k} d_i \leq \sum_{i=0}^{k} d'_i$$

for $k = 0, \cdots, n-2$ and

$$\sum_{i=0}^{n-1} d_i = \sum_{i=0}^{n-1} d'_i,$$

then π' is said to *majorize* the sequence π, denoted by

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

Note that $\pi' > \pi$ in the lexical ordering.

Lemma 2.3. [11] *Let $\pi = (d_0, \cdots d_{n-1})$ and $\pi' = (d'_0, \cdots, d'_{n-1})$ be two nonincreasing graphic degree sequences. If $\pi \triangleleft \pi'$, then there exists a series of graphic degree sequences $\pi_1, \cdots, \pi_m$ such that $\pi \triangleleft \pi_1 \triangleleft \cdots \triangleleft \pi_m \triangleleft \pi'$, where π_i and π_{i+1} differ at exactly two entries, say d_j (d'_j) and d_k (d'_k) of π_i (π_{i+1}), with $d'_j = d_j + 1$, $d'_k = d_k - 1$ and $j < k$.*

With Lemma 2.3, the following can be shown in a way similar to Theorem 2.4 in [15]. The proof is outlined for completeness.

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

$$W(T^*_\pi) > W(T^*_{\pi'})$$

where T^*_π and $T^*_{\pi'}$ are the greedy trees with degree sequences π and π' respectively.

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

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

that differ only at the j^{th} and k^{th} entries with $d'_j = d_j + 1$, $d'_k = d_k - 1$ for some $j < k$.

Let T'_π be the tree constructed from T^*_π by removing the edge vw and adding an edge uw, where u and v are the vertices corresponding to d_j and d_k respectively and w is a child of v (Figure 3).

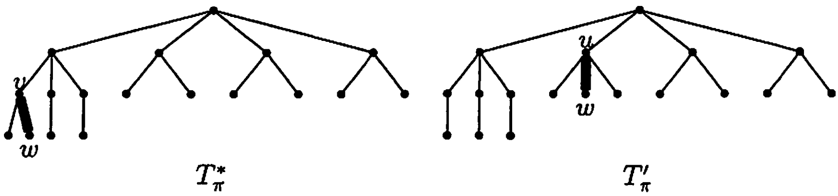


FIGURE 3. $\pi = (4, 4, 3, 3, 3, 3, 2, 2, 1, \ldots, 1)$ and $\pi' = (4, 4, 4, 3, 3, 2, 2, 2, 1, \ldots, 1)$

Let T' be the tree obtained from T^*_π after removing w and its descendants. Then the next claim follows from the structure of the greedy tree T^*_π (see, for instance, [10, 14, 15]).

Claim 2.5. *Let the path from u to v be $uu_1u_2 \ldots u_k(w)v_k \ldots v_2v_1v$ where the existence of w depends on the parity of $d(u, v)$. Let U, U_i, V, V_i denote the component containing u, u_i, v, v_i respectively after removing the edges on this path from T'. Then we have*

$$|U| \geq |V| \text{ and } |U_i| \geq |V_i|$$

for any $1 \leq i \leq k$.

Now a simple calculation shows that (see, for instance, [10])

$$W(T^*_{\pi'}) \leq W(T'_\pi) < W(T^*_\pi).$$

■

Now we are ready to consider the extremal trees with respect to $S(T)$ under additional conditions. The following results are almost immediate.

Theorem 2.6. *Among trees with given order and number of leaves, $S(T)$ is maximized by a caterpillar.*

Proof. By (3), we have

$$|V(T')| = |V(T)| - |L(T)|$$

and $W(T')$ is maximized by the path $P_{|V(T)|-|L(T)|}$. Hence, the conclusion follows. ∎

Theorem 2.7. *Among trees with given order and number of leaves, $S(T)$ is minimized by a greedy tree with degree sequence* $\left(|L(T)|, 2, \ldots, 2, \underbrace{1, \ldots, 1}_{|L(T)|\ 1\text{'s}}\right)$. *Such a tree is often called a "star-like tree" (Figure 4).*

Proof. In this case, $W(T')$ is minimized by a greedy tree with $|V(T)| - |L(T)|$ vertices and at most $|L(T)|$ leaves (since 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, it is easy to see that

$$\left(|L(T)|, \quad \underbrace{2, \ldots, 2,}_{|V(T)|-2|L(T)|-1\ 2\text{'s}} \quad \underbrace{1, \ldots, 1}_{|L(T)|\ 1\text{'s}}\right)$$

majorizes all other degree sequences. After all the leaves of T is added, we obtain a greedy tree with degree sequence

$$\left(|L(T)|, \quad \underbrace{2, \ldots, 2,}_{|V(T)|-|L(T)|-1\ 2\text{'s}} \quad \underbrace{1, \ldots, 1}_{|L(T)|\ 1\text{'s}}\right)$$

Hence, the conclusion follows. ∎

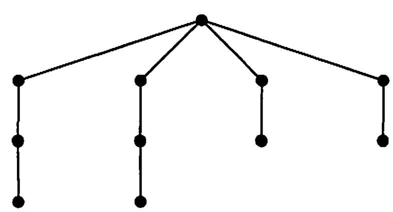


FIGURE 4. A star-like tree with $|V(T)| = 11$ and $|L(T)| = 4$

Next we consider trees with given degree sequence. It is known that $W(T)$ is maximized by a caterpillar [7] (however, to determine the extremal caterpillar with a specific degree sequence is difficult [1, 13]) and minimized by the greedy tree. The following is rather obvious.

Theorem 2.8. *Among trees with given order and degree sequence, $S(T)$ is maximized by a caterpillar and minimized by the greedy tree.*

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

To maximize $S(T)$, we once again consider T' as in (3). Since $W(T')$ is maximized by a path, $S(T)$ is maximized by a caterpillar with the given degree sequence.

To minimize $S(T)$, note that $W(T')$ is minimized by a greedy tree with the degree sequence of T'. Let the degree sequence of T be $(d_1, d_2, \ldots)$. Then the degree sequence of T' is $(d_1 - k_1, d_2 - k_2, \ldots)$ 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 $k_1 = k_2 = \ldots k_i = 0$ 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 was the greedy tree. Thus the conclusion follows from Proposition 2.4. ∎

The *complete k-ary tree* with a given maximum degree k (also called the "*good tree*" or "*Volkmann tree*" [4]) is defined in a similar way as the greedy tree, except that the vertices $v, v_1, \ldots$ take the maximum degree k until there are not enough vertices (Figure 5). As a result, the complete k-ary tree has degree sequence $(k, k, \ldots, k, m, 1, \ldots, 1)$ for some $1 < m \leq k$.

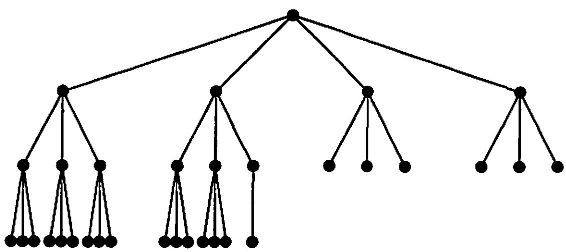


FIGURE 5. A complete 4-ary tree

For trees with given order and maximum degree, the path still maximize $S(T)$. The extremality of the complete k-ary tree follows in the same way as previous arguments, we skip the proof here.

Theorem 2.9. *Among trees with given order and maximum degree k, $S(T)$ is minimized by the complete k-ary tree.*

3. THE RELATIONS BETWEEN SOME DISTANCE-BASED INVARIANTS

When the extremal structures that maximize (minimize) different indices coincide, it is natural to ask for the existence of a nice functional relation. This is not the case in general. For example, there are trees T and T' with

$$W(T) > W(T') \text{ and } \Gamma(T) < \Gamma(T')$$

as shown in [9].

Between $W(T)$ and $S(T)$, Figure 6 shows two trees T_1 and T_2 such that

$$W(T_2) > W(T_1) \text{ and } S(T_2) < S(T_1).$$

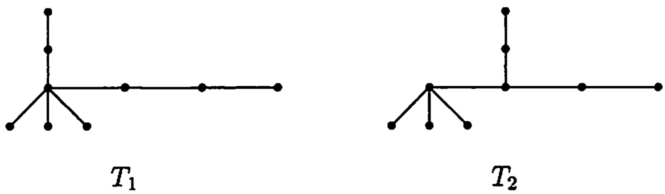


FIGURE 6. The trees T_1 and T_2

Similarly, Figure 7 shows two trees S_1 and S_2 such that

$$W(S_2) > W(S_1) \text{ and } B(S_2) < B(S_1).$$

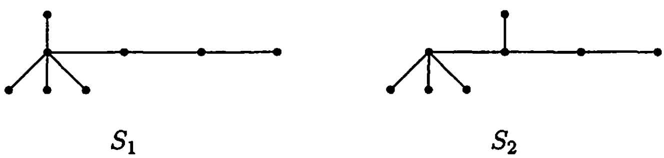


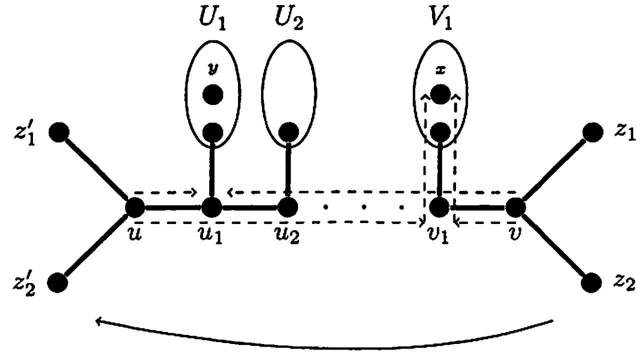
FIGURE 7. The trees S_1 and S_2

Regardless of the above examples, it is evident that $S(T)$ and $B(T)$ are closely related to $W(T)$ and $\Gamma(T)$ in some sense. As an effort to provide some insights of this problem, the following result analogous to (1) is provided below. The proof (essentially the same as the corresponding proof in [9]) is outlined below and illustrated with a binary tree in the figure below.

Proposition 3.1. *Given any two k-ary trees T_1 and T_2, we have*

$$S(T_1) - S(T_2) = \frac{1}{k-2}(\Gamma(T_1) - \Gamma(T_2)) = \frac{k-2}{(k-1)^2}(W(T_1) - W(T_2)).$$

Proof. We consider the difference $S(T_2) - S(T_1)$ for k-ary trees where T_2 is obtained from T_1 by moving the leaf-neighbors of v to u. Denote by $uu_1u_2 \dots v_2v_1v$ the path connecting u and v in T_1. Let U_i and V_i denote the corresponding components after removing this path.



In the operation from T_1 to T_2, we lose one internal vertex v and obtain one internal vertex u. The difference $S(T_2) - S(T_1)$ is then only contributed by the distances from u to other internal vertices minus the distances from v to these vertices. Choose, for instance, a non-leaf vertex $x \in V_1$ and consider the distance from v to x in T_1 and u to x in T_2 (illustrated with dotted lines above). Notice that the distances from v_1 to x cancel out in the difference and the same can be said for any non-leaf vertex $y \in U_1$, etc. Let $D_i = d(u, v_i) - d(v, v_i)$.

$$\begin{aligned} & S(T_2) - S(T_1) \\ = & \sum_i D_i \left((|\text{internal vert. of } U_i| + 1) - (|\text{internal vert. of } V_i| + 1) \right) \\ = & \sum_i D_i \frac{1}{k-2} (|L(U_i)| - |L(V_i)|) \end{aligned}$$

Comparing this to a similar expression for the difference in Γ and the expression for W from Lemma 3 of [9], the proof is finished by noting that any two different k-ary trees can be obtained from one another through a sequence of such operations.

■

We believe that a similar assertion can be proved following the same approach. In this situation, it is easier to make use of (2) and we have the following.

Corollary 3.2. *Given any two k-ary trees T_1 and T_2, we have*

$$B(T_1) - B(T_2) = \frac{k-1}{(k-2)^2} (\Gamma(T_1) - \Gamma(T_2)) = \frac{1}{k-1} (W(T_1) - W(T_2)).$$

4. Summary

Two natural distance-based graph invariants are introduced for trees. For the Spinal index, we provide a number of extremal results following some of the known techniques. For the Bartlett index, it seems that traditional approaches do not have a direct application. The characterization of extremal structures with respect to $B(T)$ seems interesting.

Proposition 3.1 and Corollary 3.2 imply a "nice" functional relation for k-ary trees between any two of the four indices under consideration. It is reasonable to consider the same question for other indices (not necessarily distance-based) among k-ary trees.

References

[1] E. Çela, N.S. Schmuck, S. Wimer, G.J. Woeginger, The Wiener maximum quadratic assignment problem, *Discrete Optim.* 8(2011) 411–416.

[2] A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: Theory and applications, *Acta Appl. Math.* 66(2001) 211–249.

[3] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, L. Volkmann, Wiener index versus maximum degree in trees, *Discrete Appl. Math.* 122(2002) 127–137.

[4] I. Gutman, A property of the Wiener number and its modifications, *Indian J. Chem.* Sec. A 36(1997) 128–132.

[5] I. Gutman, B. Furtula, M. Petrović, Terminal Wiener index, *J. Math. Chem.* 46(2)(2009) 522–531.

[6] N. Schmuck, S. Wagner, H. Wang, Greedy trees, caterpillars, and Wiener-type graph invariants, *MATCH Commun.Math.Comput.Chem.* to appear.

[7] R. Shi, The average distance of trees, *Systems Sci. Math. Sci.* 6(1) (1993) 18–24.

[8] L.A. Szekely, H. Wang, T. Wu, The sum of distances between leaves of a tree and the 'semi-regular' property, *Disc. Math.* 311(2011), 1197–1203.

[9] H. Wang, Sums of distances between vertices/leaves in k-ary trees, *Bulletin of the ICA.* 60(2010) 62–68.

[10] H. Wang, The extremal values of the Wiener index of a tree with given degree sequence, *Discrete Applied Mathematics,* 156(2008)2647–2654.

[11] W.-D.Wei, The class $\mathcal{U}(R, S)$ of (0,1) matrices, *Discrete Mathematics* 39(1982) 201–205.

[12] H. Wiener, Structural determination of paraffin boiling point, *J. Amer. Chem. Soc.* 69(1947) 17–20.

[13] X.-D. Zhang, Y. Liu and M.-X. Han, Maximum Wiener Index of Trees with Given Degree Sequence, *MATCH Commun.Math.Comput.Chem.*, 64(2010) 661–682.

[14] X.-D. Zhang, Q.-Y. Xiang, L.-Q. Xu, R.-Y. Pan, The Wiener index of trees with given degree sequences, *MATCH Commun.Math.Comput.Chem.*, 60(2008) 623–644.

[15] X.-M. Zhang, X.-D. Zhang, D. Gray, H. Wang, The number of subtrees of trees with given degree sequence, *J. Graph Theory,* to appear.

Marcus Bartlett, Department of Mathematics, Clayton State University, Morrow, GA 30260, USA
E-mail address: mbartlett2@student.clayton.edu

Elliot Krop, Department of Mathematics, Clayton State University, Morrow, GA 30260, USA
E-mail address: ElliotKrop@clayton.edu

Colton Magnant, Department of Mathematical Sciences, Georgia Southern University, Statesboro, GA 30460, USA
E-mail address: cmagnant@georgiasouthern.edu

Fedelis Mutiso, Department of Mathematical Sciences, Georgia Southern University, Statesboro, GA 30460, USA
E-mail address: fm00344@georgiasouthern.edu

Hua Wang, Department of Mathematical Sciences, Georgia Southern University, Statesboro, GA 30460, USA
E-mail address: hwang@georgiasouthern.edu