Nina Sabine SCHMUCK

The Wiener index of a graph

DIPLOMA THESIS

written to obtain the academic degree of a Diplom-Ingenieurin

Diplomstudium Technische Mathematik

Graz University of Technology

Graz University of Technology

Supervisor:
Gast-Prof. Dipl.-Ing. Dr.techn. Stephan WAGNER

Institut für Mathematik A

Graz, August 2010

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

.... date

.... (signature)

Contents

Preface

The first investigations into the Wiener index were made by Harold Wiener in 1947 who realized that there are correlations between the boiling points of paraffin and the structure of the molecules (see [23]). In particular, he mentions in his article that the boiling point t_B can be quite closely approximated by the formula

$$t_B = aw + bp + c,$$

where w is the Wiener index, p the polarity number and a, b and c are constants for a given isomeric group. Since then it has become one of the most frequently used topological indices in chemistry, as molecules are usually modelled as undirected graphs, especially trees. For example, in the drug design process, the aim is the construction of chemical compounds with certain properties, which not only depend on the chemical formula but also strongly on the molecular structure, as one can easily see when considering cocaine and scopolamine, both having the chemical formula $C_{17}H_{21}NO_4$.

Furthermore, there are many situations in communication, facility location, cryptology, architecture etc. where the Wiener index of the corresponding graph or the average distance is of great interest. One of these problems, for example, is to find a spanning tree with minimum average distance.

In the first chapter we define the Wiener index and some properties of graphs, particularly trees, that we will require later on. The second chapter deals with a variety of formulas for computing the Wiener index of trees. Some of those formulas can also be applied to connected graphs.

Since calculating the Wiener index of a graph can be computationally expensive, we give some cheaply computable lower and upper bounds for the Wiener index – given certain graph properties – in chapter 3. In particular, we focus on trees with further constraints, such as a given maximum degree or degree sequence.

Finally, in chapter 4, we consider the inverse problem – Which numbers are Wiener indices? – for a number of classes of graphs and trees.

I want to thank all who helped me during my study of mathematics and the writing of my diploma thesis, especially Gast-Prof. Dr. Stephan Wagner and Ao.Univ.-Prof. Dr. Clemens Heuberger. Finally, I also want to thank my family for their support and Bernd and Doris Haug for their linguistic advice.

Graz, August 2010

Nina Sabine Schmuck

Chapter 1

Introduction - Basic notations and definitions

Throughout the whole diploma thesis, all graphs will be finite, simple, connected and undirected, and in most cases we will consider trees.

Definition 1.1. Let G be a graph with vertex set $V(G)$ and edge set $E(G)$. The *distance* $d_G(u,v)$ between two vertices u, $v \in V(G)$ is the minimum number of edges on a path in G between u and v.

Definition 1.2. Let G be as before. The *Wiener index* $W(G)$ of G is defined by

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

and the *average distance* $\mu(G)$ between the vertices of G by

$$\mu(G) = \frac{W(G)}{\binom{|V(G)|}{2}}.$$

Definition 1.3. Let G be as before. The *distance* $d_G(v)$ *of a vertex* v is the sum of all distances between v and all other vertices of G.

Thus, one can define the Wiener index also in a slightly different way:

$$W(G) = \frac{1}{2} \sum_{v \in V(G)} d_G(v)\,,$$

where $\frac{1}{2}$ compensates for the fact that each path between u and v is counted in $d_G(u)$ as well as in $d_G(v)$.

Definition 1.4. Let G be a graph. The *diameter* $d(G)$ of G is defined as

$$d(G) = \min_{\{u,v\} \subseteq V(G)} d_G(u,v).$$

Definition 1.5. The *degree* $\deg_G(v)$ of a vertex $v \in V(G)$ is the number of edges incident to v.

The *degree sequence* of G is a vector $(\deg_G(v_1), \deg_G(v_2), \ldots, \deg_G(v_n))$ with $\deg_G(v_1) \geq \deg_G(v_2) \geq \cdots \geq \deg_G(v_n)$ and $n = |V(G)|$.

Furthermore we will need the following definitions:

Definition 1.6. Let T be a tree. A vertex $v \in V(T)$ is called *branching point of T*, if $\deg_T(v) \geq 3$. If $\deg_T(v) = 1$, the vertex v is named *leaf of T*.

The *path* with n vertices, of which exactly 2 are leaves, is written as P_n, and the *star* with exactly $n-1$ leaves and 1 branching point is denoted by S_n.

Remark 1.1. It is easy to see that every tree on n vertices has at least 2 leaves and at most $\frac{n-2}{2}$ branching points.

Remark 1.2. As we will need the Wiener index of both the path and the star on several occasions, we compute their Wiener index in advance:

$$W(S_n) = \underbrace{n-1}_{\text{branching point to leaves}} + \underbrace{2\sum_{i=1}^{n-2} i}_{\text{between all pairs of leaves}} = (n-1)^2,$$

$$W(P_n) = \sum_{i=1}^{n}\sum_{j=1}^{i-1} j = \sum_{i=1}^{n}\binom{i}{2} = \binom{n+1}{3}.$$

The last equation can be easily shown by induction.

Definition 1.7. Let T be a tree. A *segment* S of T is a path-subtree whose terminal vertices are branching points or leaves and internal vertices v have degree $\deg_T(v) = 2$. The *length of a segment* S is equal to the number of edges in S and is denoted by l_S. The set of all vertices being terminal vertices of a segment is named by $SP(T)$. Moreover let S^* be a subpath of S containing l_S vertices, i.e. $S^* = S\backslash\{v\}$ where v is a terminal vertex of S, and let S^0 be S without both its terminal vertices.

Remark 1.3. Since each edge of a tree T on n vertices is used in exactly one segment, it is clear that

$$\sum_{S \text{ seg. of } T} l_S = n-1.$$

To illustrate the above definitions let us examine the following two short examples:

Example 1.4. Let G be the graph shown in Figure 1.1. The distance e.g. between v_4 and v_6 is $d_G(v_4, v_6) = 2$ using the path (v_4, v_7, v_6), and the distance e.g. of v_1 is $d_G(v_1) = 1+2+3+4+2+3+4 = 19$. The Wiener index of G can be computed as follows:

$$W(G) = \sum_{i=1}^{7}\sum_{j=i+1}^{8} d_G(v_i, v_j)$$

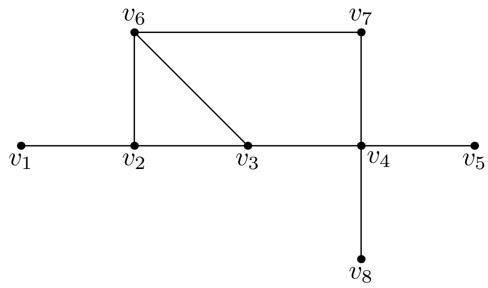


Figure 1.1: A graph G used in Example 1.4.

$$\begin{aligned} &= (1+2+3+4+2+3+4) + (1+2+3+1+2+3)+ \\ &\quad + (1+2+1+2+2) + (1+2+1+1)+ \\ &\quad + (3+2+2) + (1+3) + 2 \\ &= 57. \end{aligned}$$

Thus the average distance of G is

$$\mu(G) = \frac{57}{\binom{8}{2}} = \frac{57}{28} \approx 2.0357.$$

Furthermore the degree sequence of G is $(4, 3, 3, 3, 2, 1, 1, 1)$.

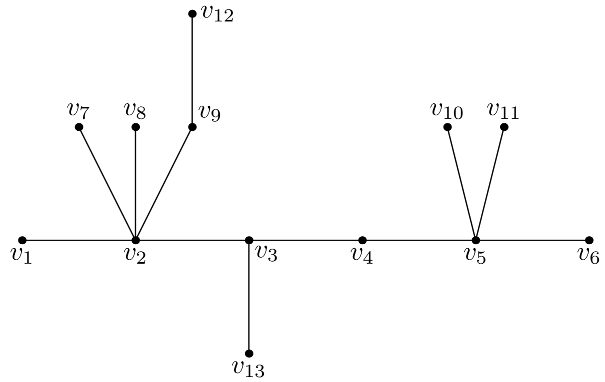


Figure 1.2: A tree T used in Example 1.5.

Example 1.5. Now let us consider the tree T as shown in Figure 1.2. For example v_1 is a leaf and v_2 a branching point, but notice that v_9 is neither a leaf nor a branching point since $\deg_T(v_9) = 2$.

A segment of T is e.g. $S_1 = (v_2, v_9, v_{12})$ with length $l_{S_1} = 2$. Furthermore S_1^* can either be the path (v_2, v_9) or the path (v_9, v_{12}), and S_1^0 is the path only containing the vertex v_9. Another segment would be $S_2 = (v_3, v_{13})$ with S_2^* either (v_3) or (v_{13}) and S_2^0 empty.

Chapter 2

Different ways to compute the Wiener index of trees

As the path and therefore the distance between two vertices of a tree is unique, the Wiener index of a tree is much easier to compute than that of an arbitrary graph. In the following, we will show different formulas for computing the Wiener index, in the first part direct ones and in the second part recursive ones that require certain characteristics of the trees but in case their requirements are satisfied may make the calculation of the Wiener index easier by far.

2.1 Direct formulas

The first formula we are going to show is a very basic one and was found by H. Wiener in 1947 (see [23]). While the definition of the Wiener index puts its stress on how far one has to go from each vertex to reach all other vertices, this formula counts how often one has to pass each edge.

Definition 2.1. Let $e = (u, v) \in E(T)$ be an edge of the tree T. The subtrees T_u and T_v are defined as the connected components of T containing u and v, respectively. The order of the subtrees is denoted by $n_u(e) = |V(T_u)|$ and $n_v(e) = |V(T_v)|$.

Theorem 2.1. *Let T be a tree. Then*

$$W(T) = \sum_{e=(u,v)\in E(T)} n_u(e)n_v(e). \tag{2.1}$$

Proof. As T is a tree, the unique path between a vertex $x \in V(T_u)$ and a vertex $y \in V(T_v)$ must contain e. If x and y are chosen in a different way, e is not part of the path between them. Therefore $n_u(e)n_v(e)$ is exactly the number of times how often e belongs to a path between two vertices of T. Then the sum of $n_u(e)n_v(e)$ over all edges of T must be the Wiener index of T. □

Remark 2.2. Dobrynin and Gutman give another proof of Theorem 2.1 in a more general way in [6]. In addition to that they show for what types of graphs equation (2.1) holds and that for all other cases the right-hand side of equation 2.1, which is denoted as the new graph invariant W^* called *Szeged index*, is greater than the corresponding Wiener index. To be able to define W^* one has to generalize the definition of T_u and T_v first, which we are going to show in the following definition to give an idea of this closely related graph invariant before continuing with our original topic.

Definition 2.2. Let $e = (u, v) \in E(G)$ be an edge of the graph G. The sets $B_u(e)$ and $B_v(e)$ of vertices of G are defined as

$$B_u(e) = \{x \in V(G) : d_G(x, u) < d_G(x, v)\}$$

$$B_v(e) = \{y \in V(G) : d_G(y, v) < d_G(y, u)\}.$$

The cardinalities of the sets are denoted by $n_u(e) = |B_u(e)|$ and $n_v(e) = |B_v(e)|$.

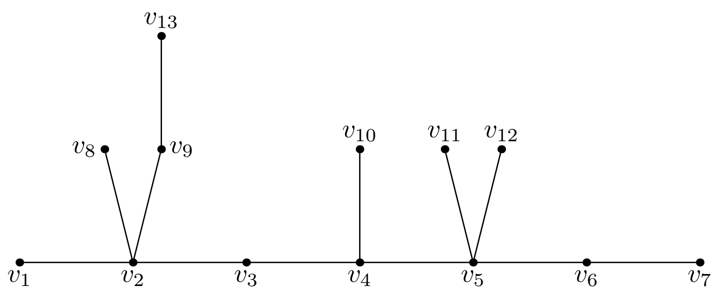


Figure 2.1: An arbitrary tree T.

Example 2.3. Since each edge $e = (u, v)$ of a tree is a bridge, it is obvious that $n_u(e) + n_v(e) = n$ and thus one only needs to count the vertices lying on one side of e. Therefore, by using Theorem 2.1, we can easily compute the Wiener index of the tree T of order $n = 13$ shown in Figure 2.1:

$$W(T) = 7 \cdot 1 \cdot 12 + 2 \cdot 2 \cdot 11 + 2 \cdot 5 \cdot 8 + 1 \cdot 6 \cdot 7 = 250.$$

Within the same paper, Dobrynin and Gutman also show the following:

Theorem 2.4. *Let T be a tree on n vertices. Then the Wiener index can be computed in the following way:*

$$W(T) = \frac{1}{4}\left[n^2(n-1) - \sum_{(u,v)\in E(T)} [d_T(v) - d_T(u)]^2\right].$$

Proof. As mentioned in Example 2.3, $n = n_u(e) + n_v(e)$ for all edges $e = (u, v)$. Further we obtain

$$\begin{aligned}
d_T(v) - d_T(u) &= \left(\sum_{x \in B_u(e)} d_T(v, x) + \sum_{y \in B_v(e)} d_T(v, y) \right) \\
&\quad - \left(\sum_{x \in B_u(e)} d_T(u, x) + \sum_{y \in B_v(e)} d_T(u, y) \right) \\
&= \sum_{x \in B_u(e)} (d_T(v, x) - d_T(u, x)) - \sum_{y \in B_v(e)} (d_T(u, y) - d_T(v, y)) \\
&= \sum_{x \in B_u(e)} 1 - \sum_{y \in B_v(e)} 1 \\
&= n_u(e) - n_v(e).
\end{aligned}$$

Thus together we obtain $2n_u(e) = n + (d_T(v) - d_T(u))$ and $2n_v(e) = n - (d_T(v) - d_T(u))$. Substituting this into equation (2.1) we get

$$\begin{aligned}
W(T) &= \sum_{(u,v) \in E(T)} \frac{1}{2}(n + d_T(v) - d_T(u)) \frac{1}{2}(n - d_T(v) + d_T(u)) \\
&= \frac{1}{4} \sum_{(u,v) \in E(T)} \left[n^2 - (d_T(v) - d_T(u))^2 \right] \\
&= \frac{1}{4} \left[n^2(n-1) - \sum_{(u,v) \in E(T)} (d_T(v) - d_T(u))^2 \right],
\end{aligned}$$

which completes the proof. □

Example 2.5. Let T again be the tree in Figure 2.1. To apply Theorem 2.4 we make use of the fact that

$$d_T(v) - d_T(u) = n_u(e) - n_v(e)$$

for all edges $e = (u, v)$ as we showed within the proof of Theorem 2.4. This simplifies the calculation by far and we obtain

$$W(T) = \frac{1}{4}[13^2 \cdot 12 - (7 \cdot 11^2 + 2 \cdot 9^2 + 2 \cdot 3^2 + 1^2)] = 250.$$

Corollary 2.6. *Let T be a tree on n vertices. Then*

$$W(T) = \frac{1}{4} \left[n(n-1) + \sum_{v \in V(T)} \deg_T(v) d_T(v) \right].$$

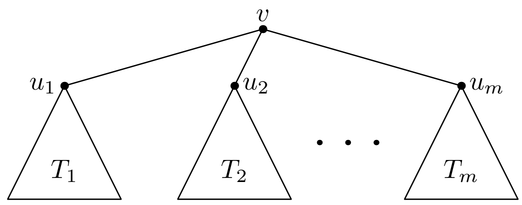


Figure 2.2: Tree with branching point v and subtrees T_i, where $m = \deg_T(v)$

Proof. Let $N(v)$ be the set of all neighbours of the vertex v in T. It is obvious that $|N(v)| = \deg_T(v)$. Then we can rewrite the sum in Theorem 2.4 as follows

$$\begin{aligned}
W(T) &= \frac{1}{4}\left[n^2(n-1) - \sum_{(u,v)\in E(T)} [d_T(v) - d_T(u)]^2\right] \\
&= \frac{1}{4}\left[n^2(n-1) - \frac{1}{2}\sum_{v\in V(T)}\sum_{u\in N(v)} (d_T(v)^2 - 2d_T(v)d_T(u) + d_T(u)^2)\right] \\
&= \frac{1}{4}\Bigg[n^2(n-1) - \frac{1}{2}\sum_{v\in V(T)}\Big(d_T(v)^2\deg_T(v) \\
&\quad - 2\sum_{u\in N(v)} d_T(v)d_T(u) + d_T(v)^2\deg_T(v)\Big)\Bigg] \\
&= \frac{1}{4}\left[n^2(n-1) - \sum_{v\in V(T)} d_T(v)\Big(d_T(v)\deg_T(v) - \sum_{u\in N(v)} d_T(u)\Big)\right].
\end{aligned}$$

To compute $d_T(u_i)$ with $u_i \in N(v)$ we choose the path $u_i \to v \to x$ for every $x \in V(T)$. Now we counted two edges too many for each vertex in T_i, where T_i is defined as the connected component of T including u_i after deleting v (as shown in Figure 2.2). Therefore we get

$$d_T(u_i) = n + d_T(v) - 2|T_i|,$$

and altogether we obtain

$$\sum_{u\in N(v)} d_T(u) = \deg_T(v)n + d_T(v)\deg_T(v) - 2(n-1).$$

Substituting this into $W(T)$ leads to

$$\begin{aligned}
W(T) &= \frac{1}{4}\left[n^2(n-1) - \sum_{v\in V(T)} (2(n-1)d_T(v) - nd_T(v)\deg_T(v))\right] \\
&= \frac{1}{4}\left[n^2(n-1) - 4(n-1)W(T) + n\sum_{v\in V(T)} d_T(v)\deg_T(v)\right].
\end{aligned}$$

Solving the equation for $W(T)$ leads to the desired statement. $\square$

Example 2.7. Now we apply Corollary 2.6 to the tree T of Figure 2.1:

$$W(T) = \frac{1}{4}[13 \cdot 12 + (1 \cdot 42 + 4 \cdot 31 + 2 \cdot 28 + 3 \cdot 27 + 4 \cdot 30 + 2 \cdot 39 + 1 \cdot 50 + 1 \cdot 42 + 2 \cdot 40 + 1 \cdot 38 + 1 \cdot 41 + 1 \cdot 41 + 1 \cdot 51)] = 250.$$

Remark 2.8. Examples 2.5 and 2.7 illustrate quite well that the formulas used there are more of theoretical than of practical interest.

2.1.1 Branching points, segments and the Wiener index

Another very important formula is given by Doyle and Graver in [7]. But at first, we need the following definition:

Definition 2.3. Let G be a connected graph. The vertices v_1, v_2, $v_3 \in V(G)$ are called *collinear* if there exists an ordering of the three vertices such that

$$d_G(v_i, v_j) + d_G(v_j, v_k) = d_G(v_i, v_k).$$

The number of 3-subsets of $V(G)$ which are not collinear is denoted by $\tau(G)$.

Theorem 2.9 (Doyle-Graver formula). *Let T be a tree of order n. Then*

$$W(T) = \binom{n+1}{3} - \sum_{v \in V(T)} \sum_{1 \le i < j < k \le \deg_T(v)} |V(T_i)|\, |V(T_j)|\, |V(T_k)| \tag{2.2}$$

with T_l defined as in the proof of Corollary 2.6.

Proof. Let C be the set of all collinear 3-subsets of $V(T)$. As the path between two vertices u and v is unique, the only vertices w being collinear with u and v, such that $d_T(u, w) + d_T(w, v) = d_T(u, v)$, are the vertices on the path between u and v. This means that for each pair u and v there exists exactly $d_T(u, v) - 1$ vertices that are collinear with them in this manner. Therefore we get

$$|C| = \sum_{\{u,v\} \subseteq V(T)} (d_T(u, v) - 1) = W(T) - \binom{n}{2}.$$

Since all vertices u, v, w can be either collinear or non-collinear, we obtain

$$|C| + \tau(T) = \binom{n}{3}.$$

Combining these two formulas leads to

$$W(T) = \binom{n}{3} + \binom{n}{2} - \tau(T) = \binom{n+1}{3} - \tau(T). \tag{2.3}$$

Now, considering three non-collinear vertices u_1, u_2, u_3 it is obvious that there must exist a vertex v such that v lies on the paths between each pair of u_1, u_2, u_3. Counting the number of non-collinear 3-subsets of $V(T)$ with v on their paths we get exactly $\sum_{1\le i<j<k\le \deg_T(v)} |V(T_i)|\,|V(T_j)|\,|V(T_k)|$. Thus

$$\tau(T) = \sum_{v\in V(T)} \sum_{1\le i<j<k\le \deg_T(v)} |V(T_i)|\,|V(T_j)|\,|V(T_k)|.$$

Substituting this into equation (2.3) completes the proof. □

Remark 2.10. Notice that the first summation in equation (2.2) actually goes just over all branching points of T. Furthermore notice that the first term in equation (2.2) is exactly the Wiener index of the path of length n.

Example 2.11. The tree T shown in Figure 2.1 has the three branching points v_2, v_4 and v_5. Since $\deg_T(v_4) = 3$, the vertex v_4 contributes just one addend to the Doyle-Graver formula, whereas both v_2 and v_5, have degree 4, which leads to $\binom{4}{3} = 4$ different combinations of their subtrees. This leads to the Wiener index

$$\begin{aligned} W(T) = \binom{14}{3} - [&\underbrace{1\cdot 1\cdot 2 + 1\cdot 1\cdot 8 + 1\cdot 2\cdot 8 + 1\cdot 2\cdot 8}_{\text{obtained from } v_2} + \underbrace{6\cdot 1\cdot 5}_{\text{obtained from } v_4} \\ &+ \underbrace{8\cdot 1\cdot 1 + 8\cdot 1\cdot 2 + 8\cdot 1\cdot 2 + 1\cdot 1\cdot 2}_{\text{obtained from } v_5}] = 250, \end{aligned}$$

which is once again the same number.

On the following pages some further formulas based on either branching points or segments are examined (see [4]).

Theorem 2.12. *Let T be a tree on n vertices. Then*

$$W(T) = \sum_{S \text{ seg. of } T} n_1(S) n_{l_S+1}(S) l_S + \frac{1}{6} \sum_{S \text{ seg. of } T} l_S(l_S-1)(3n-2l_S+1) \tag{2.4}$$

where $n_1(S)$ and $n_{l_S+1}(S)$ are the number of vertices of the two connected components obtained by deleting all internal vertices of S and the corresponding edges (compare Definition 2.1).

Proof. In order to prove this theorem we use formula (2.1) and rewrite it in terms of segments. Let $S = (v_1, v_2, \ldots, v_{l_S}, v_{l_S+1})$ be a segment of T and $e_i = (v_i, v_{i+1})$. Then we obtain $n_{v_i}(e_i) = n_1(S) + (i-1)$ and $n_{v_{i+1}}(e_i) = n_{l_S+1}(S) + (l_S - i)$ for $i \in \{1, 2, \ldots, l_S\}$ and clearly $n_1(S) + n_{l_S+1}(S) + l_S - 1 = n$. Thus the contribution of the edges of S to $W(T)$ is

$$\sum_{i=1}^{l_S} n_{v_i}(e_i) n_{v_{i+1}}(e_i) = \sum_{i=1}^{l_S} (n_1(S) + i - 1)(n_{l_S+1}(S) + l_S - i)$$

$$
\begin{aligned}
&= \sum_{i=1}^{l_S} [n_1(S)n_{l_S+1}(S) + (n_1(S)-1)l_S - n_{l_S+1}(S) \\
&\quad + (n_{l_S+1}(S) - n_1(S) + l_S + 1)i - i^2] \\
&= l_S n_1(S)n_{l_S+1}(S) + (n_1(S)-1)l_S^2 - n_{l_S+1}(S)l_S \\
&\quad + (n_{l_S+1}(S) - n_1(S) + l_S + 1)\frac{l_S(l_S+1)}{2} - \frac{l_S(l_S+1)(2l_S+1)}{6} \\
&= l_S n_1(S)n_{l_S+1}(S) + \frac{1}{2}(n_1(S) + n_{l_S+1}(S) + l_S - 1)l_S^2 \\
&\quad + \frac{1}{2}(-n_1(S) - n_{l_S+1}(S) - l_S + 1 + 2l_S)l_S - \frac{1}{6}(2l_S^2 + 3l_S + 1)l_S \\
&= l_S n_1(S)n_{l_S+1}(S) + \frac{1}{2}nl_S^2 + \frac{1}{2}(-n + 2l_S)l_S - \frac{1}{6}(2l_S^2 + 3l_S + 1)l_S \\
&= l_S n_1(S)n_{l_S+1}(S) + \frac{1}{6}l_S(3nl_S - 3n - 2l_S^2 + 3l_S - 1).
\end{aligned}
$$

Summing this over all segments of T leads to the desired equation. □

Example 2.13. The tree shown in Figure 2.1 has nine segments. As the segments $S_1 = (v_1, v_2)$, $S_2 = (v_2, v_8)$, $S_3 = (v_4, v_{10})$, $S_4 = (v_5, v_{11})$ and $S_5 = (v_5, v_{12})$ are equivalent according to their length l_{S_i} and the number of vertices $n_1(S_i)$ and $n_{l_{S_i}+1}(S_i)$, we get $n_1(S_i)n_{l_{S_i}+1}(S_i)l_{S_i} = 12$ for $i = 1, \ldots, 5$. In the same manner we obtain for both segments $S_6 = (v_2, v_9, v_{13})$ and $S_7 = (v_5, v_6, v_7)$ that the term $n_1(S)n_{l_S+1}(S)l_S$ is 22. Furthermore we have $n_1(S_8)n_{l_{S_8}+1}(S_8)l_{S_8} = 40$ for $S_8 = (v_4, v_5)$ and $n_1(S_9)n_{l_{S_9}+1}(S_9)l_{S_9} = 70$ for $S_9 = (v_2, v_3, v_4)$. Thus according to equation (2.4) we obtain

$$W(T) = 5 \cdot 12 + 2 \cdot 22 + 40 + 70 + \frac{1}{6} \cdot 3 \cdot [2 \cdot 1 \cdot (3 \cdot 13 - 2 \cdot 2 + 1)] = 250.$$

Theorem 2.14. *Let T be a tree of order n. Then the Wiener index can be computed by*

$$W(T) = \frac{1}{12}\left[(3n^2+1)(n-1) - 3\sum_{S \text{ seg. of } T} \frac{1}{l_S}[d_T(v_1) - d_T(v_{l_S+1})]^2 - \sum_{S \text{ seg. of } T} l_S^3\right]$$

with v_1 and v_{l_S+1} being the terminal vertices of S.

Proof. In order to apply equation (2.4) we have to further investigate the product $n_1(S)n_{l_S+1}(S)$. Let $T_1(S)$ and $T_{l_S+1}(S)$ with $n_1(S)$ and $n_{l_S+1}(S)$ vertices, respectively, be the two connected subtrees obtained by deleting all inner vertices of S. Thus we get

$$
\begin{aligned}
d_T(v_1) - d_T(v_{l_S+1}) &= \sum_{x \in T_1(S)} (d_T(v_1, x) - d_T(v_{l_S+1}, x)) \\
&\quad + \sum_{y \in T_{l_S+1}(S)} (d_T(v_1, x) - d_T(v_{l_S+1}, x))
\end{aligned}
$$

$$= \sum_{x \in T_1(S)} (-l_S) + \sum_{y \in T_{l_S+1}(S)} l_S$$
$$= l_S[n_{l_S+1}(S) - n_1(S)].$$

Furthermore we know that $n_1(S) + n_{l_S+1}(S) = n - l_S + 1$. Together we obtain

$$\begin{aligned} n_1(S)n_{l_S+1}(S) &= \frac{1}{4}([n_1(S) + n_{l_S+1}(S)]^2 - [n_1(S) - n_{l_S+1}(S)]^2) \\ &= \frac{1}{4}\left[(n - l_S + 1)^2 - \frac{[d_T(v_1) - d_T(v_{l_S+1})]^2}{l_S^2}\right] \end{aligned}$$

and substituting this into equation (2.4) leads to

$$\begin{aligned} W(T) &= \frac{1}{4} \sum_{S \text{ seg. of } T} (n - l_S + 1)^2 l_S - \frac{1}{4} \sum_{S \text{ seg. of } T} \frac{1}{l_S}[d_T(v_1) - d_T(v_{l_S+1})]^2 \\ &\quad + \frac{1}{6} \sum_{S \text{ seg. of } T} l_S(l_S - 1)(3n - 2l_S + 1) \\ &= \frac{1}{12}\left[\sum_S (3n^2 + 1)l_S - \sum_S l_S^3 - 3\sum_S \frac{1}{l_S}[d_T(v_1) - d_T(v_{l_S+1})]^2\right] \\ &= \frac{1}{12}\left[(3n^2 + 1)(n - 1) - 3\sum_S \frac{1}{l_S}[d_T(v_1) - d_T(v_{l_S+1})]^2 - \sum_S l_S^3\right]. \end{aligned}$$

□

Example 2.15. Since $d_T(v_1) - d_T(v_{l_S+1}) = (n_{l_S+1}(S) - n_1(S))l_S$ with notation as in Theorem 2.14, we obtain for the tree in Figure 2.1 by using Theorem 2.14

$$W(T) = \frac{1}{12}[(3 \cdot 13^2 + 1)12 - 3(5 \cdot 11^2 + 2 \cdot 10^2 \cdot 2 + 3^2 + 2^2 \cdot 2) - (6 \cdot 1^3 + 3 \cdot 2^3)] = 250.$$

Next we are going to rewrite Theorem 2.6 in terms of segments and generalized stars. Therefore we need the following definition and lemma:

Definition 2.4. A *generalized star* associated with a vertex $v \in V(T)$, T a tree, consists of v and all segments beginning at v, and q_v denotes its number of edges.

Remember that S^* is the segment S without one, and S^0 the segment S without both terminal vertices. Furthermore we define $SP(T)$ to be the set of all terminal vertices of segments of the tree T.

Lemma 2.16. *Let $S = (v_1, v_2, \ldots, v_{l_S+1})$ be a segment of the tree T on n vertices. Then*

$$\sum_{i=2}^{l_S} d_T(v_i) = \frac{1}{2}(l_S - 1)[d_T(v_1) + d_T(v_{l_S+1})] - \frac{1}{6} l_S(l_S^2 - 1)$$

Proof. Let $T_1(S)$, $T_{l_S+1}(S)$, $n_1(S)$ and $n_{l_S+1}(S)$ be defined as before. Then the sum of all distances of v_i for $i \in \{2, 3, \dots, l_S\}$ is

$$\begin{aligned}
d_T(v_i) = & \sum_{x \in V(T_1(S))} [d_T(v_i, v_1) + d_T(v_1, x)] \\
& + \sum_{y \in V(T_{l_S+1}(S))} [d_T(v_i, v_{l_S+1}) + d_T(v_{l_S+1}, y)] + d_{S^0}(v_i) \\
= & \sum_{x \in V(T_1(S))} d_T(v_1, x) + \sum_{y \in V(T_{l_S+1}(S))} d_T(v_{l_S+1}, y) \\
& + n_1(S)(i-1) + n_{l_S+1}(S)(l_S - i + 1) + d_{S^0}(v_i).
\end{aligned}$$

For the two terminal vertices we obtain

$$\begin{aligned}
d_T(v_1) &= \sum_{x \in V(T_1(S))} d_T(v_1, x) + \sum_{y \in V(T_{l_S+1}(S))} [d_T(v_1, v_{l_S+1}) + d_T(v_{l_S+1}, y)] + d_{S^*}(v_1) \\
&= \sum_{x \in V(T_1(S))} d_T(v_1, x) + \sum_{y \in V(T_{l_S+1}(S))} d_T(v_{l_S+1}, y) + n_{l_S+1}(S) l_S + \binom{l_S}{2}
\end{aligned}$$

and analogously

$$d_T(v_{l_S+1}) = \sum_{x \in V(T_1(S))} d_T(v_1, x) + \sum_{y \in V(T_{l_S+1}(S))} d_T(v_{l_S+1}, y) + n_1(S) l_S + \binom{l_S}{2}.$$

Therefore we get

$$\begin{aligned}
d_T(v_1) + d_T(v_{l_S+1}) = & \; 2 \sum_{x \in V(T_1(S))} d_T(v_1, x) + 2 \sum_{y \in V(T_{l_S+1}(S))} d_T(v_{l_S+1}, y) \\
& + [n_1(S) + n_{l_S+1}(S)] l_S + l_S(l_S + 1)
\end{aligned}$$

which leads to

$$\begin{aligned}
\sum_{x \in V(T_1(S))} d_T(v_1, x) + \sum_{y \in V(T_{l_S+1}(S))} d_T(v_{l_S+1}, y) = & \; \frac{1}{2}[d_T(v_1) + d_T(v_{l_S+1}) - l_S(l_S - 1) \\
& - [n_1(S) + n_{l_S+1}(S)] l_S]
\end{aligned}$$

Finally we can calculate

$$\begin{aligned}
\sum_{i=2}^{l_S} d_T(v_i) = & \; \frac{1}{2} \sum_{i=2}^{l_S} [d_T(v_1) + d_T(v_{l_S+1}) - l_S(l_S - 1) - [n_1(S) + n_{l_S+1}(S)] l_S] \\
& + n_1(S) \sum_{i=2}^{l_S} (i-1) + n_{l_S+1}(S) \sum_{i=2}^{l_S} (l_S - i + 1) + \sum_{i=2}^{l_S} d_{S^0}(v_i)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}(l_S-1)[d_T(v_1)+d_T(v_{l_S+1})] - \frac{1}{2}l_S(l_S-1)^2 + 2W(S^0)\\
&\quad - \frac{1}{2}l_S(l_S-1)[n_1(S)+n_{l_S+1}(S)] + [n_1(S)+n_{l_S+1}(S)]\binom{l_S}{2}\\
&= \frac{1}{2}(l_S-1)[d_T(v_1)+d_T(v_{l_S+1})] - \frac{1}{2}l_S(l_S-1)^2 + 2\binom{l_S}{3}\\
&= \frac{1}{2}(l_S-1)[d_T(v_1)+d_T(v_{l_S+1})] - \frac{1}{6}l_S(l_S-1)(3l_S-3-2l_S+4)\\
&= \frac{1}{2}(l_S-1)[d_T(v_1)+d_T(v_{l_S+1})] - \frac{1}{6}l_S(l_S^2-1).
\end{aligned}
$$

□

Theorem 2.17. *Let T be a tree on n vertices. Then*

$$W(T) = \frac{1}{12}\left[(3n+1)(n-1) + 3\sum_{v\in SP(T)} q_v d_T(v) - \sum_{S \text{ seg. of } T} l_S^3\right]$$

Proof. As a vertex v can either have degree equal to 2 (then it is an inner vertex of a segment S) or degree greater than 2 (then it is a terminal vertex of S) we can rewrite the formula in Corollary 2.6 as follows:

$$
\begin{aligned}
W(T) &= \frac{1}{4}\left[n(n-1) + \sum_{v\in SP(T)} \deg_T(v)d_T(v) + \sum_{S\in T}\sum_{i=2}^{l_S} 2d_T(v_i)\right]\\
&= \frac{1}{4}\Bigg[n(n-1) + \sum_{v\in SP(T)} \deg_T(v)d_T(v)\\
&\quad + 2\sum_{S\in T}\Big(\frac{1}{2}(l_S-1)[d_T(v_1)+d_T(v_{l_S+1})] - \frac{1}{6}l_S(l_S^2-1)\Big)\Bigg]\\
&= \frac{1}{12}\Bigg[3n(n-1) + 3\sum_{v\in SP(T)} \deg_T(v)d_T(v)\\
&\quad + 3\sum_{S\in T}(l_S-1)[d_T(v_1)+d_T(v_{l_S+1})] - \sum_{S\in T} l_S^3 + \sum_{S\in T} l_S\Bigg]\\
&= \frac{1}{12}\Bigg[3n(n-1) + 3\sum_{v\in SP(T)} \deg_T(v)d_T(v)\\
&\quad + 3\sum_{v\in SP(T)}(q_v-\deg_T(v))d_T(v) - \sum_{S\in T} l_S^3 + (n-1)\Bigg]\\
&= \frac{1}{12}\left[(3n+1)(n-1) + 3\sum_{v\in SP(T)} q_v d_T(v) - \sum_{S\in T} l_S^3\right].
\end{aligned}
$$

□

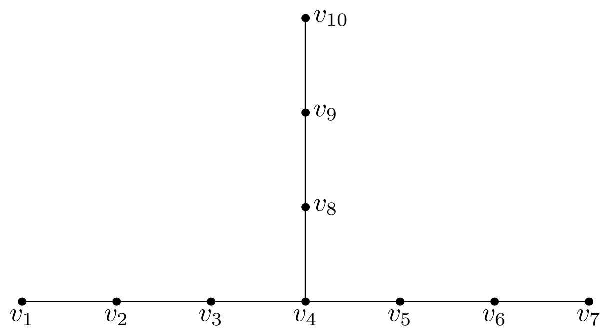


Figure 2.3

Example 2.18. Let us consider the tree T shown in Figure 2.3. As T consists of three segments and because of symmetry, we only need to compute the distances of two vertices in T. Thus it makes sense to use Theorem 2.17 to calculate the Wiener index of T and we obtain

$$W(T) = \frac{1}{12}[(3 \cdot 10 + 1)9 + 3(3 \cdot 3 \cdot 36 + 9 \cdot 18) - 3 \cdot 3^3] = 138.$$

2.1.2 Laplacian Eigenvalues and their influence on the Wiener index

A completely different way to compute the Wiener index of a tree is by using the eigenvalues of its Laplacian matrix. The formula for computing the Wiener index, which I am going to present in this subsection, was published independently in several papers around 1990, but I will confine myself here to mentioning the proof given by Merris in [17].

Definition 2.5. Let G be a graph. The *Laplacian matrix* L_G is given by

$$L(G) = D(G) - A(G)$$

where $D(G)$ is the diagonal matrix of the vertex degrees and $A(G)$ the adjacency matrix.

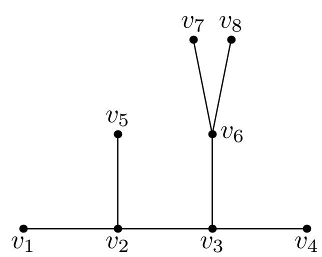


Figure 2.4

Example 2.19. Let T be the tree shown in Figure 2.4. Its Laplacian matrix is

$$L(T) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 3 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0-1 & 0 & 1 \end{pmatrix}.$$

A very important statement involving the Laplacian matrix is the celebrated matrix-tree theorem, also called Kirchhoff's theorem after Gustav Kirchhoff who implicitly mentioned it in 1847 (see [13]).

Theorem 2.20 (Matrix-tree theorem)**.** *Let $L(G)$ be the Laplacian matrix of the graph G on n vertices and $L_{i,j}$ denote the submatrix of L formed by crossing out row i and column j. Then the number of spanning trees $\tau(G)$ is obtained by*

$$\tau(G) = (-1)^{i+j} \det(L_{i,j})$$

for all i, $j \in \{1, 2, \ldots, n\}$.

For a proof of the matrix-tree theorem see e.g. [15].

Lemma 2.21. *Let G be an oriented graph with $|E(G)| = m$ and $Q = Q(G) = (q_{ij})$ its vertex-edge incidence matrix defined in the following manner*

$$q_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is the positive end of } e_j, \\ -1 & \text{if } v_i \text{ is the negative end of } e_j, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$L(G) = QQ^t.$$

Proof. Let $R = (r_{ij})$ be the matrix on the right side. Then $r_{ij} = \sum_{a=1}^{m} q_{ia}q_{ja}$.

Case $i \neq j$: The product $q_{ia}q_{ja}$ is not 0, if v_i and v_j are both terminal vertices of e_a. Because of the orientation of G we get $q_{ia}q_{ja} = -1$ and thus $r_{ij} = -a_{ij}$.

Case $i = j$: Here we obtain $r_{ii} = \sum_{a=1}^{m} q_{ia}^2$ with $q_{ia}^2 = 1$ if v_i is a terminal vertex of e_a and $q_{ia}^2 = 0$ otherwise. This means $r_{ii} = \deg_G(v_i)$. □

Remark 2.22. Note that although the entries of Q depend on the chosen orientation, the Laplacian matrix $L(G)$ is independent of it. $L(G)$ is unique up to permutation which also means that the eigenvalues of $L(G)$ do not depend on the orientation of G or the order of its vertices.

Unlike $L(G)$ its "edge version" $K(G) = Q^tQ$ depends on the orientation for the signs of its off-diagonal entries. Due to the singular value decomposition of Q, the eigenvalues not equal to 0 of $K(G)$ and $L(G)$ are the same.

Denote the eigenvalues of $L(G)$ by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then, as the sum over the entries of each row and column is 0, $\lambda_n = 0$ and, following from the matrix-tree theorem, $\lambda_{n-1} > 0$ if and only if G is connected.

Lemma 2.23. *Let T be a tree on n vertices and $K(T)$ as before. Then*

$$\det(K(T)) = n.$$

Proof. According to Remark 2.22 we obtain

$$\det(K(T)) = \prod_{i=1}^{n-1} \lambda_i$$

with λ_i, $i \in \{1, 2, \ldots, n-1\}$, the eigenvalues of $L(T)$ not equal to 0. Considering the characteristic polynomial of $L(T)$ it is easy to see that the coefficient coeff_λ of λ is

$$\text{coeff}_\lambda = (-1)^{n-1} \prod_{i=1}^{n-1} \lambda_i.$$

On the other hand, it is well known (see e.g. [14]) that

$$\text{coeff}_\lambda = (-1)^{n-1} \sum_{i=1}^{n} \det(L_{i,i}).$$

Thus we get

$$\det(K(T)) = \sum_{i=1}^{n} \det(L_{i,i}) = n$$

since $L_{i,i} = 1$ due to the matrix-tree theorem. □

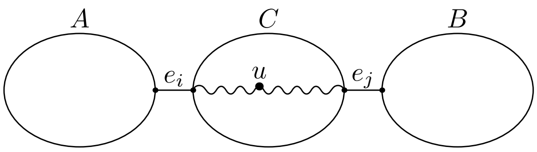


Figure 2.5

Definition 2.6. Let T be a tree and e_i, e_j two edges of T. Furthermore let $T' = (V(T), E(T) \backslash \{e_i, e_j\})$ be the forest with three components A, B and C, where A and B are on opposite sides of e_i and e_j such that a path in T from any vertex of A to any vertex of B contains e_i and e_j (see Figure 2.5). Then we define $n(e_i, e_j) = |A|\,|B|$. Analogously we use $n(e_i) = n_u(e_i)n_v(e_i)$ for $e_i = (u, v)$.

A vertex u of C is called *between* e_i and e_j if every path from a vertex of A to a vertex of B passes through u. The number of vertices between e_i and e_j is denoted by s_{ij}.

Lemma 2.24. *Let T be a tree on n vertices and orient T such that $K(T)$ is not negative. Let $X = (x_{ij})$ be the adjugate of $K(T)$. Then*

$$x_{ij} = \begin{cases} n(e_i) & \textit{if } i = j \\ (-1)^{s_{ij}} n(e_i, e_j) & \textit{otherwise.} \end{cases}$$

Proof. Denote by $K_{i,j}$ the submatrix of $K(T)$ obtained by deleting row i and column j, corresponding, respectively, to edge e_i and e_j and, analogously, by $Q_{pq,i}$ the submatrix of Q after deleting the rows p and q, corresponding to the vertices v_p and v_q, and column i, corresponding to edge e_i. Then

$$\begin{aligned} x_{ij} &= (-1)^{i+j} \det K_{i,j} \\ &= (-1)^{i+j} \sum_{p=1}^{n-1} \sum_{q=p+1}^{n} \det Q^t_{i,pq} \det Q_{pq,j} \\ &= (-1)^{i+j} \sum_{p=1}^{n-1} \sum_{q=p+1}^{n} \det Q_{pq,i} \det Q_{pq,j}. \end{aligned} \tag{2.5}$$

The second equality we get by using the Binet-Cauchy formula as $K_{i,j} = Q^t_{i,\_} Q_{\_,j}$. Thus we have to compute the determinant for any $(n-2)$-square submatrix of Q. Remember that in each column of Q there are just two non-zero elements. If we delete a row p of Q which means removing vertex v_p but not any edges from T, there is exactly one permutation π such that $\prod_{\substack{l=1 \\ l \neq p}}^{n} q_{l\pi(l)} \neq 0$. This can easily be seen, as choosing $q_{l\pi(l)}$ means to assign edge $e_{\pi(l)}$ to its terminal vertex v_l. Analogously we obtain for $Q_{pq,i}$ that there exists exactly one permutation if and only if the edge e_i lies on the path between v_p and v_q. Therefore we get

$$\det Q_{pq,i} = \begin{cases} \pm 1 & \text{if } e_i \text{ is between } v_p \text{ and } v_q, \\ 0 & \text{otherwise.} \end{cases}$$

Considering the case $i = j$ we obtain

$$\begin{aligned} x_{ii} &= (-1)^{2i} \sum_{p=1}^{n-1} \sum_{q=p+1}^{n} \det Q_{pq,i} \det Q_{pq,i} \\ &= \sum_{p=1}^{n-1} \sum_{q=p+1}^{n} (\det Q_{pq,i})^2 \\ &= n(e_i) \end{aligned}$$

since x_{ii} counts the pairs p, q, $p < q$, such that e_i lies on the path between v_p and v_q, which means that x_{ii} is the product of the numbers of vertices on each side of e_i.

Now we assume $i \neq j$ and, without loss of generality, we can further assume $i < j$. In order for a product $\det Q_{pq,i} \det Q_{pq,j}$ to be non-zero of course both factors must be non-zero, which is only possible if the vertices v_p and v_q lie on opposite sides of e_i as well as on opposite sides of e_j. Thus $|x_{ij}| \leq n(e_i, e_j)$, and equality is attained only if the signs of all summands are the same. Recall that we are assuming the orientation of the edges has been chosen such that $K(T)$ is entry-wise non-negative, i.e.

$$k_{ab} = \sum_{c=1}^{n} q_{ca} q_{cb} \geq 0 \quad \forall a, b.$$

It is obvious that $q_{ca}q_{cb} \neq 0$ if and only if e_a and e_b have v_c as their common terminal vertex. Since T is a tree, each pair of edges can have at most one common terminal vertex, i.e. at most one summand can be non-zero. Therefore $k_{ab} = 0$ if e_a and e_b have no common terminal vertex, and $k_{ab} = 1$ otherwise. In the second case both edges are oriented the same way according to their common terminal vertex. This implies that all entries of a row of Q have the same sign. Hence the only difference between the permutations making a contribution to the computation of $\det Q_{pq,i}$ and $\det Q_{pq,j}$ is that for the non-zero elements of column i and column j one has to choose another non-zero element of the same row. Thus the only factor to influence the sign of $\det Q_{pq,i} \det Q_{pq,j}$ are the signs of the permutations. Let the permutations be denoted by $\pi_{pq,i}$ and $\pi_{pq,j}$ and their signs $\mathrm{sgn}(pq, i)$ and $\mathrm{sgn}(pq, j)$. It is easily seen that the signs depend on both the arbitrary ordering of the vertices and of the edges, but their product only depends on how the edges are ordered since K is independent of the numbering of the vertices. Thus, we may assume that the vertices between e_i and e_j are consecutively numbered, say $k+1$, $k+2$, …, $k+s$ with $s = s_{ij}$, and v_{k+1} is a terminal vertex of e_i and v_{k+s} a terminal vertex of e_j. The only difference between the two permutations $\pi_{pq,i}$ and $\pi_{pq,j}$ is the assignment of the edges lying on the path between e_i and e_j to their terminal edges. If e_i is deleted, we obtain the assignment $v_{k+s} \leftrightarrow e_j$ and $v_{k+l} \leftrightarrow (v_{k+l}, v_{k+l+1})$ for $l = 1, \ldots, s-1$. On the other hand, if e_j is deleted, we get the assignment $v_{k+1} \leftrightarrow e_i$ and $v_{k+l} \leftrightarrow (v_{k+l-1}, v_{k+l})$ for $l = 2, \ldots, s$. This means, for getting the permutation $\pi_{pq,i}$ by changing permutation $\pi_{pq,j}$, we perform the following matrix operations: move column i past $j-i-1$ columns and row $k+1$ past $s-1$ rows. Therefore

$$\mathrm{sgn}(pq, i) = (-1)^{j-i-1}(-1)^{s-1}\mathrm{sgn}(pq, j)$$

and thus every non-zero product in equation (2.5) has the same sign, namely $(-1)^{s_{ij}+j-i}$. Hence, the absolute value of x_{ij} is $n(e_i, e_j)$ and altogether we obtain

$$\begin{aligned} x_{ij} &= (-1)^{i+j}(-1)^{s_{ij}+j-i} n(e_i, e_j) \\ &= (-1)^{s_{ij}} n(e_i, e_j). \end{aligned}$$

□

Now we have all tools needed to prove the following main statement:

Theorem 2.25. *Let T be a tree on n vertices and $\lambda_1 \geq \cdots \geq \lambda_{n-1} > \lambda_n = 0$ the eigenvalues of the corresponding Laplacian matrix. Then the Wiener index can be computed as*

$$W(T) = n \sum_{i=1}^{n-1} \frac{1}{\lambda_i}.$$

Proof. We choose an orientation of T such that $K(T)$ is not negative. According to Lemma 2.24 the trace of the adjugate X of $K(T)$ is

$$\operatorname{tr}(X) = \sum_{i=1}^{n-1} x_{ii} = \sum_{e \in E(T)} n(e_i).$$

As shown in Theorem 2.1, the last sum is equal to the Wiener index of T. On the other hand we know that the trace of a matrix can also be computed by using its eigenvalues. For all λ_i, $i = 1, \ldots, n-1$, we obtain that $n\frac{1}{\lambda_i}$ is an eigenvalue of $X = nK^{-1}(T)$. Thus

$$\operatorname{tr}(X) = \sum_{i=1}^{n-1} n \frac{1}{\lambda_i},$$

which completes the proof. □

Example 2.26. Let T again be the tree in Figure 2.4. Then the non-zero eigenvalues of $L(T)$ are approximately

$$\begin{aligned}
\lambda_1 &= 4.81361,\\
\lambda_2 &= 3.73205,\\
\lambda_3 &= 2.52932,\\
\lambda_4 &= \lambda_5 = 1,\\
\lambda_6 &= 0.657077,\\
\lambda_7 &= 0.267949
\end{aligned}$$

and according to Theorem 2.25 we obtain $W(T) = 8 \sum_{i=1}^{7} \frac{1}{\lambda_i} = 65$.

2.2 Recursive formulas

Up to now we have just seen explicit formulas for computing the Wiener index. In some cases it may be easier not to compute the Wiener index of the tree itself but of special subtrees and get the Wiener index of the whole tree as a combination of these subtrees, e.g. when the tree T is obtained by connecting several copies of the tree T' at a vertex u, always using the same vertex $v' \in V(T')$. Thus, in the following some recursive formulas will be shown.

Maybe the first idea that occurs if one wants to calculate the Wiener index recursively is to take a tree, delete a leaf and compute the Wiener index of the remaining subtree.

Theorem 2.27. *Let T be a tree on $n \geq 2$ vertices and $v \in V(T)$ a leaf of T. Furthermore let $(u, v) \in E(T)$ and $T' = T - v$ be the subgraph of T after deleting v. Then*

$$W(T) = W(T') + d_{T'}(u) + n - 1.$$

Proof. Let x and y be two vertices of T. If $v \neq x$ and $v \neq y$, the distance between x and y does not change after deleting v. Therefore the sum of all these distances is the Wiener index of T'. If one of the two vertices is v, w.l.o.g. $v = x$, then $d_T(x, y) = d_{T'}(u, y) + 1$. Thus the sum of all $n-1$ pairs $\{x, y\}$ equals $d_{T'}(u) + n - 1$, which completes the proof. □

Example 2.28. Consider a tree T with n vertices, obtained by taking S_{n-1} and connecting a further vertex v to a leaf u of S_{n-1}. Using Theorem 2.27 we easily compute

$$\begin{aligned} W(T) &= W(S_{n-1}) + d_{S_{n-1}}(u) + n - 1 \\ &= (n-2)^2 + (n-3)2 + 1 + n - 1 \\ &= n^2 - n - 2. \end{aligned}$$

Since the fact, that T is a tree, was just used for counting the number of vertex pairs $\{v, y\}$, it is obvious that Theorem 2.27 can also be generalized for connected graphs. We state this formula in the next theorem because of completeness.

Theorem 2.29. *Let G be a connected graph and $v \in V(G)$ a leaf. Besides let $(v, u) \in E(G)$ and $G' = G - v$ be the subgraph of G after deleting v. Then*

$$W(G) = W(G') + d_{G'}(u) + |V(G')|.$$

A generalization of Theorem 2.27, where not only a leaf but an arbitrary vertex can be deleted, is given in [3]:

Theorem 2.30. *Let T be a tree of order $n \geq 2$ as shown in Figure 2.2. Then*

$$W(T) = \sum_{i=1}^{m} [W(T_i) + (n - |V(T_i)|) d_{T_i}(u_i) - |V(T_i)|^2] + n(n-1).$$

Proof. To compute $d_T(x, y)$ with $x \in T_i$ fixed, we have to consider two cases, according to whether $y \in T_i$ or not. If $y \in T_i$, we have $d_T(x, y) = d_{T_i}(x, y)$ and the sum of all such pairs is equal to $W(T_i)$.

In case $y \in T_j$, $i \neq j$, we obtain

$$d_T(x, y) = d_{T_i}(x, u_i) + d_T(u_i, v) + d_T(u_j, v) + d_{T_j}(y, u_j).$$

Thus we get

$$W(T) = \sum_{i=1}^{m} \left[W(T_i) + \frac{1}{2} \sum_{x \in T_i} \sum_{\substack{y \in T_j \\ j \neq i}} (d_{T_i}(x, u_i) + d_T(u_i, v) + d_T(u_j, v) + d_{T_j}(y, u_j)) \right]$$

$$
\begin{aligned}
&= \sum_{i=1}^{m} \left[W(T_i) + \sum_{x \in T_i} \sum_{\substack{y \in T_j \\ j \neq i}} (d_{T_i}(x, u_i) + d_T(u_i, v)) \right] \\
&= \sum_{i=1}^{m} \left[W(T_i) + \sum_{x \in T_i} (d_{T_i}(x, u_i) + d_T(u_i, v))(n - |V(T_i)|) \right] \\
&= \sum_{i=1}^{m} \left[W(T_i) + (n - |V(T_i)|)(d_{T_i}(u_i) + |V(T_i)|) \right] \\
&= \sum_{i=1}^{m} \left[W(T_i) + (n - |V(T_i)|) d_{T_i}(u_i) - |V(T_i)|^2) \right] + n \sum_{i=1}^{m} |V(T_i)| \\
&= \sum_{i=1}^{m} \left[W(T_i) + (n - |V(T_i)|) d_{T_i}(u_i) - |V(T_i)|^2) \right] + n(n-1).
\end{aligned}
$$

□

Example 2.31. Let T be a tree on n vertices, obtained by taking the path P_{l+1} and attaching $n-l-1$ vertices to one of its leaves v. To compute $W(T)$ we apply Theorem 2.30 with vertex v as the separating point and therefore with $n-l-1$ single vertices and one path of length l as the subtrees. Thus we obtain

$$
\begin{aligned}
W(T) &= \sum_{i=1}^{n-l-1} [0 + (n-1)0 - 1^2] \binom{l+1}{3} + (n-l)\binom{l}{2} - l^2 + n(n-1) \\
&= \binom{l+1}{3} + (n-l)\binom{l}{2} - (n-1-l) - l^2 + n(n-1) \\
&= \binom{l}{3} + (n-l-1)\binom{l}{2} + (n-1)^2.
\end{aligned}
$$

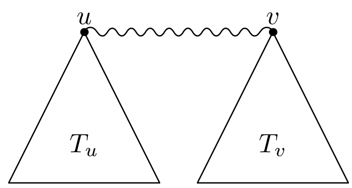


Figure 2.6: The trees T_u and T_v connected by a path of k new vertices.

In [5] the following formula for the Wiener index of a tree which arises from two trees by connecting them by a path can be found:

Theorem 2.32. *Let T_u and T_v be two trees with $n_u = |V(T_u)|$ and $n_v = |V(T_v)|$ and $u \in V(T_u)$ and $v \in V(T_v)$ two vertices. T arises from T_u and T_v by connecting u and v by* *a path on k new vertices (see Figure 2.6). Then*

$$\begin{aligned}W(T) &= W(T_u) + W(T_v) + (n_u + k)d_{T_v}(v) + (n_v + k)d_{T_u}(u) + (k+1)n_u n_v\\&\quad + \frac{1}{2}(k^2 + k)(n_u + n_v) + \frac{1}{6}(k^3 - k).\end{aligned}$$

Proof. Let x, y be two vertices of T. Then we distinguish the following cases:

Case 1: $x,\ y \in V(T_i)$, $i = u,\ v$. It is obvious that $d_T(x, y) = d_{T_i}(x, y)$.

Case 2: $x \in V(T_u)$ and $y \in V(T_v)$, then $d_T(x, y) = d_{T_u}(x, u) + \underbrace{d_T(u, v)}_{=k+1} + d_{T_v}(v, y)$.

Case 3: x and y are both new vertices. The contribution of all such vertices is the Wiener index of a path of length $k - 1$.

Case 4: $x \in V(T_i)$, $i = u,\ v$, and y is one of the new vertices. In this case we obtain $d_T(x, y) = d_{T_i}(x, i) + d_T(i, y)$.

Therefore we get

$$\begin{aligned}W(T) = &\sum_{\{x,y\}\subseteq V(T_u)} d_{T_u}(x, y) + \sum_{\{x,y\}\subseteq V(T_v)} d_{T_v}(x, y)\\&+ \sum_{x\in V(T_u)} \sum_{y\in V(T_v)} (d_{T_u}(x, u) + k + 1 + d_{T_v}(v, y)) + \binom{k+1}{3}\\&+ \sum_{\substack{y \text{ new}\\ \text{vertex}}} \left(\sum_{x\in V(T_u)} (d_{T_u}(x, u) + d_T(u, y)) + \sum_{x\in V(T_v)} (d_{T_v}(x, v) + d_T(v, y)) \right)\\= &\ W(T_u) + W(T_v) + d_{T_u}(u)n_v + n_u n_v(k + 1) + d_{T_v}(v)n_u + \frac{1}{6}(k^3 - k)\\&+ kd_{T_u}(u) + (n_u + n_v)\sum_{i=1}^{k} i + kd_{T_v}(v)\\= &\ W(T_u) + W(T_v) + (n_u + k)d_{T_v}(v) + (n_v + k)d_{T_u}(u) + (k + 1)n_u n_v\\&+ \frac{1}{2}(k^2 + k)(n_u + n_v) + \frac{1}{6}(k^3 - k),\end{aligned}$$

which completes the proof. □

Example 2.33. Let T be the tree arising from S_{14} and S_{19} by connecting the branching point u of S_{14} with the branching point v of S_{19} by a path on 12 new vertices. It is easy to see that $d_{S_{14}}(u) = 13$ and $d_{S_{19}}(v) = 18$. Furthermore we know that $W(S_l) = (l - 1)^2$. Thus we obtain

$$\begin{aligned}W(T) &= 13^2 + 18^2 + (14 + 12)18 + (19 + 12)13 + 13 \cdot 14 \cdot 19\\&\quad + \frac{1}{2}(12^2 + 12)(19 + 14) + \frac{1}{6}(12^3 - 12) = 7682.\end{aligned}$$

Remark 2.34. Since nowhere in the proof of Theorem 2.32 the fact that T_u and T_v are trees has been used, it is obvious that the formula remains the same for T_u and T_v arbitrary graphs.

Corollary 2.35. *Let T_u and T_v be two trees of orders n_u and n_v, respectively, and vertices $u \in V(T_u)$, $v \in V(T_v)$. If T arises from T_u and T_v by connecting u and v by an edge,*

$$W(T) = W(T_u) + W(T_v) + n_u d_{T_v}(v) + n_v d_{T_u}(u) + n_u n_v.$$

Proof. Evidently this formula is exactly the formula of Theorem 2.32 with $k = 0$. □

In the same paper they suggest a different way of connecting two trees: To take a path of each tree containing no branching points and to identify these vertices in the same order as they appear within the paths. This leads to the following theorem:

Theorem 2.36. *Let T_1 and T_2 be two trees with $n_1 = |V(T_1)|$ and $n_2 = |V(T_2)|$. Furthermore let $p_1 = (u_1, u_2, \dots, u_k)$ be a path in T_1 and $p_2 = (v_1, v_2, \dots, v_k)$ a path in T_2, both without branching points. Then the Wiener index of the tree T, which is obtained by identifying u_i and v_i, $i = 1, \dots, k$, can be computed as*

$$\begin{aligned} W(T) = W(T_1) &+ W(T_2) + (n_1 - k)d_{T_2}(v_1) + (n_2 - k)d_{T_1}(u_1) \\ &+ 2(k-1)[n_{u_k}(p_1) + n_{v_k}(p_2) - n_{u_k}(p_1)n_{v_k}(p_2)] \\ &- \frac{1}{2}k(k-1)(n_1+n_2) + \frac{1}{6}(k-1)(5k^2 - k - 12), \end{aligned}$$

where $n_{u_k}(p_1)$ is the number of vertices in the connected component of T_1 containing u_k after deleting all edges of p_1 and $n_{v_k}(p_2)$ is defined analogously.

Proof. Let T_1', T_1'', T_2' and T_2'' be defined as shown in Figure 2.7. Then $|V(T_1'')| = n_{u_k}(p_1) - 1$, $|V(T_2'')| = n_{v_k}(p_2) - 1$ and with that, we trivially get $|V(T_1')| = n_1 - n_{u_k}(p_1) - k + 1$, $|V(T_2')| = n_2 - n_{v_k}(p_2) - k + 1$.

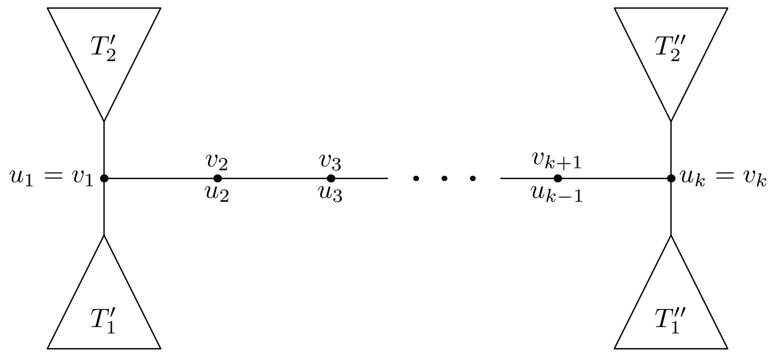


Figure 2.7

To calculate the Wiener index of T we once again distinguish between some different types of vertex pairs:

Of course all the distances between vertices of T_1 keep being the same, for which reason the contribution of these vertices is $W(T_1)$. Under the same considerations, the contribution of vertex pairs of T_2 is $W(T_2)$. As the vertices of p_1 and p_2 are identified we have to subtract all distances between these vertices once (they are counted once in $W(T_1)$ and once in $W(T_2)$), which is $\binom{k+1}{3}$.

Next we calculate the distances within T_2 between vertices of T_1' and vertices of $T_2 - p_2$. Clearly each vertex of T_1' contributes $d_{T_2}(v_1) - \binom{k}{2}$ where the second term compensates for the fact that $d_{T_2}(v_1)$ also includes the distances to the vertices of p_2. In the same manner we obtain $d_{T_1}(u_1) - \binom{k}{2}$ for the distance within T_1 between a vertex of T_2' and all vertices of $T_1 - p_1$.

Using also v_1 as the point of reference we get $d_{T_2}(v_1) - \binom{k}{2} - (k-1)(n_{v_k}(p_2) - 1)$ as the distances within T_2 between a vertex of T_1'' and all vertices of $T_2 - p_2$. The third term occurs because for all vertices of T_2'' the (not needed) path between v_1 and v_k is counted in $d_{T_2}(v_1)$. Analogously we obtain $d_{T_1}(u_1) - \binom{k}{2} - (k-1)(n_{u_k}(p_1) - 1)$ for the distances within T_2 between a vertex of T_2'' and all vertices of $T_1 - p_1$.

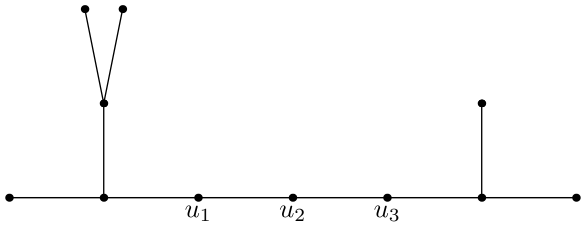
Combining all cases we get

$$\begin{aligned}
W(T) &= W(T_1) + W(T_2) - \binom{k+1}{3} + (n_1 - n_{u_k}(p_1) - k + 1)\left(d_{T_2}(v_1) - \binom{k}{2}\right) \\
&\quad + (n_2 - n_{v_k}(p_2) - k + 1)\left(d_{T_1}(u_1) - \binom{k}{2}\right) \\
&\quad + (n_{u_k}(p_1) - 1)\left(d_{T_2}(v_1) - \binom{k}{2} - (k-1)(n_{v_k}(p_2) - 1)\right) \\
&\quad + (n_{v_k}(p_2) - 1)\left(d_{T_1}(u_1) - \binom{k}{2} - (k-1)(n_{u_k}(p_1) - 1)\right) \\
&= W(T_1) + W(T_2) - \frac{1}{6}(k-1)(k^2+k) + (n_1 - k)d_{T_2}(v_1) + (n_2 - k)d_{T_1}(u_1) \\
&\quad - \frac{1}{2}k(k-1)(n_1 + n_2 - 2k) - 2(k-1)(n_{u_k}(p_1) - 1)(n_{v_k}(p_2) - 1) \\
&= W(T_1) + W(T_2) + (n_1 - k)d_{T_2}(v_1) + (n_2 - k)d_{T_1}(u_1) \\
&\quad + 2(k-1)[n_{u_k}(p_1) + n_{v_k}(p_2) - n_{u_k}(p_1)n_{v_k}(p_2)] \\
&\quad - \frac{1}{2}k(k-1)(n_1 + n_2 - 2k) + \frac{1}{6}(k-1)\underbrace{(-k^2 - k + 6k^2 - 12)}_{=5k^2-k-12},
\end{aligned}$$

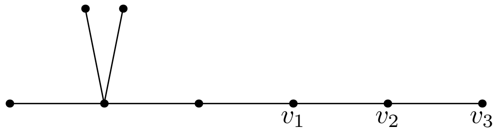
which completes the proof. □

Example 2.37. Let T, T_1 and T_2 be the trees shown in Figure 2.8. Furthermore let $p_1 = (u_1, u_2, u_3)$ and $p_2 = (v_1, v_2, v_3)$. Then we have $n_1 = 11$, $n_2 = 8$, $k = 3$, $d_{T_1}(u_1) = 25$, $d_{T_2}(v_1) = 15$, $n_{u_3}(p_1) = 4$ and $n_{v_3}(p_2) = 1$. Besides we easily compute $W(T_1) = 186$ and $W(T_2) = 71$ by using e.g. the Doyle-Graver formula. According Theorem 2.36 we obtain

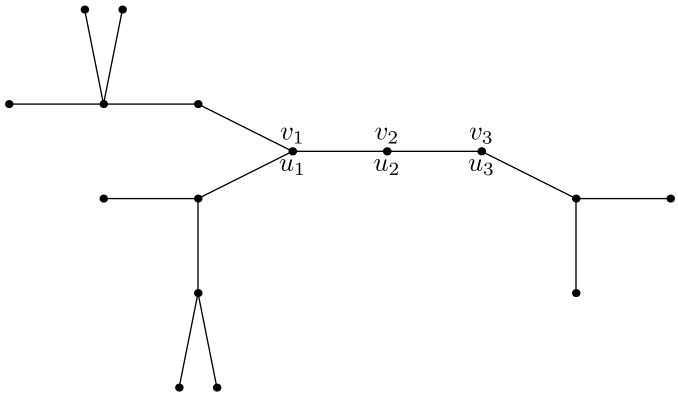
$$\begin{aligned}
W(T) &= 186 + 71 + (11-3)15 + (8-3)25 + 2 \cdot 2(4+1-4) \\
&\quad - \frac{1}{2} \cdot 3 \cdot 2(11+8) + \frac{1}{6} \cdot 2(5 \cdot 3^2 - 3 - 12) = 459.
\end{aligned}$$



(a) Tree T_1 with path $p_1 = (u_1, u_2, u_3)$.



(b) Tree T_2 with path $p_2 = (v_1, v_2, v_3)$.



(c) Tree T which arose from T_1 and T_2 by identifying p_1 and p_2.

Figure 2.8

Corollary 2.38. *Let T_1 and T_2 be two trees on n_1 and n_2 vertices, respectively, and vertices $u \in V(T_1)$ and $v \in V(T_2)$. Furthermore let T be the tree arising from T_1 and T_2 by identifying u and v. Then*

$$W(T) = W(T_1) + W(T_2) + (n_1 - 1)d_{T_2}(v_2) + (n_2 - 1)d_{T_1}(v_1).$$

Proof. Obviously we get this formula by using Theorem 2.36 with $k = 1$. □

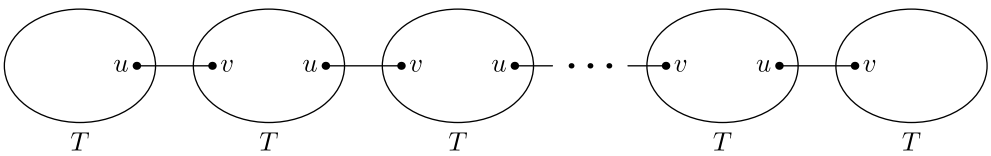


Figure 2.9: Fasciagraph F with generating tree T.

The above formulas just deal with a tree obtained by connecting only two trees. In the following theorem, which can be found in [18], we are considering a tree F obtained by connecting copies of the same graph T in a chain such that the vertex u of one copy of T is linked to the vertex v of the next copy of T by an edge. An example of a so called *fasciagraph* F is shown in Figure 2.9.

Theorem 2.39. *Let F be the fasciagraph formed by m copies of a tree T on n vertices, $m, n \geq 1$, and u, $v \in V(T)$ be the vertices by which the copies of T are linked. Then*

$$\begin{aligned} W(F) = mW(T) &+ \frac{1}{2}nm(m-1)[d_T(u) + d_T(v)] \\ &+ \frac{1}{6}n^2m(m-1)[(m-2)d_T(u,v) + m + 1]. \end{aligned}$$

Proof. There are two different ways of choosing a pair of vertices: either they both lie in the same copy of T or they lie in different ones. If they are in the same copy of T, the contribution of all these pairs is $mW(T)$, which can easily be seen.

So let x be a vertex of the i-th copy of T and y of the $(i+j)$-th one. For better readability we will write T_k for the k-th copy of T. Then the distance between x and y is $d_F(x,y) = d_T(x,u) + j + (j-1)d_T(v,u) + d_T(v,y)$, where the second term on the right side stands for the number of edges connecting all copies of T on the way from T_i to T_{i+j}. The third term describes the number of edges one has to pass within the other $j-1$ copies of T on the way from x to y.

Together, we obtain

$$\begin{aligned} W(T) &= mW(T) + \sum_{i=1}^{m-1}\sum_{j=1}^{m-i}\sum_{x\in T_i}\sum_{y\in T_{i+j}} [d_T(x,u) + j + (j-1)d_T(v,u) + d_T(v,y)] \\ &= mW(T) + n\sum_{i=1}^{m-1}\sum_{j=1}^{m-i}[d_T(u) + d_T(v)] + n^2\sum_{i=1}^{m-1}\sum_{j=1}^{m-i} j \\ &\quad + n^2\sum_{i=1}^{m-1}\sum_{j=1}^{m-i}(j-1)d_T(u,v) \\ &= mW(T) + n[d_T(u) + d_T(v)]\sum_{i=1}^{m-1}(m-i) + n^2\sum_{i=1}^{m-1}\frac{(m-i+1)(m-i)}{2} \\ &\quad + n^2 d_T(u,v)\sum_{i=1}^{m-1}\frac{(m-i)(m-i-1)}{2} \\ &= mW(T) + n\frac{m(m-1)}{2}[d_T(u) + d_T(v)] + n^2\sum_{i=1}^{m-1}\binom{i+1}{2} \\ &\quad + n^2 d_T(u,v)\sum_{i=1}^{m-2}\binom{i+1}{2} \\ &= mW(T) + \frac{1}{2}nm(m-1)[d_T(u) + d_T(v)] + n^2\binom{m+1}{3} \\ &\quad + n^2 d_T(u,v)\binom{m}{3} \\ &= mW(T) + \frac{1}{2}nm(m-1)[d_T(u) + d_T(v)] \end{aligned}$$

$$+\frac{1}{6}n^2m(m-1)[(m-2)d_T(u,v)+m+1].$$

□

Remark 2.40. Notice that nowhere in the proof of Theorem 2.39 the fact that T is a tree was used. Thus the formula holds also for T an arbitrary connected graph.

Example 2.41. To illustrate Theorem 2.39 we consider the fasciagraph $F_{n,m}$ obtained by connecting m copies of the star S_n such that both u and v are the vertex with degree $\deg_{S_n}(u) = n-1$. Thus we obtain

$$\begin{aligned} W(F_{n,m}) &= m(n-1)^2 + \frac{1}{2}nm(m-1)2(n-1) + \frac{1}{6}n^2m(m-1)(m+1) \\ &= m[(n-1)(nm-1) + \frac{1}{6}n^2(m^2-1)]. \end{aligned}$$

2.2.1 The Wiener index of a thorn tree

A completely different concept of reducing the Wiener index of a tree T_1 to the Wiener index of another tree T_2 is to add some new leaves to T_2:

Definition 2.7. Let T be a tree of order n. Then T^* is called *thorn tree* of T if T^* arises form T by attaching n_i new vertices to the vertex v_i of T, $i = 1, 2, \ldots, n$.

Remark 2.42. It is obvious that the number of vertices of T^* is $n^* = n + \sum_{i=1}^{n} n_i$ and $\deg_{T^*}(v_i) = \deg_T(v_i) + n_i$.

Furthermore notice that neither the thorn tree of a given tree nor the tree from which the thorn tree has arisen is unique.

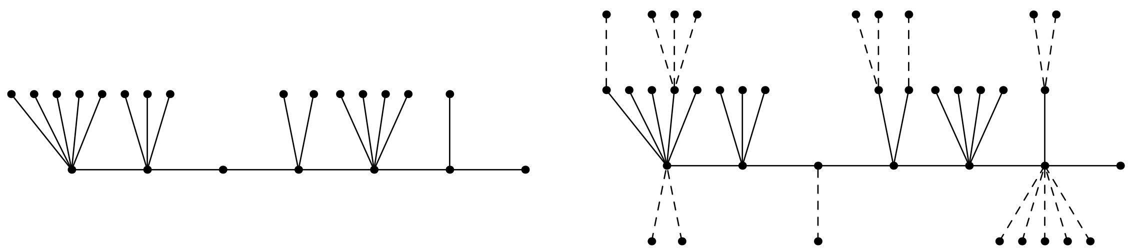


Figure 2.10: A tree and its thorn tree.

Example 2.43. An example of a tree T and one of its possible thorn trees T^* is given in Figure 2.10, where the dashed edges indicate the new ones of T^*. As one can see not all vertices of T have to be connected with a new vertex.

The following formula for computing the Wiener index of the thorn tree T^* by using the Wiener index of the corresponding tree T was given by Gutman in 1998 [10]:

Theorem 2.44. *Let T be a tree on n vertices and T^* its thorn tree. Then*

$$\begin{aligned} W(T^*) &= W(T) + \sum_{1\le i<j\le n} (n_i+n_j)d_T(v_i,v_j) + \sum_{1\le i<j\le n} n_i n_j d_T(v_i,v_j) \\ &\quad + \left(\sum_{i=1}^{n} n_i\right)^2 + (n-1)\sum_{i=1}^{n} n_i \\ &= \sum_{1\le i<j\le n} (n_i+1)(n_j+1)d_T(v_i,v_j) + \left(\sum_{i=1}^{n} n_i\right)^2 + (n-1)\sum_{i=1}^{n} n_i \end{aligned}$$

with n_i the number of new vertices connected to v_i.

Proof. Let $V(T) = \{v_1, v_2, \ldots, v_n\}$. For each pair of vertices x, $y \in V(T^*)$ we distinguish between four cases:

Case 1: $x \in V(T)$ and $y \in V(T)$. It is obvious that $d_{T^*}(x,y) = d_T(x,y)$ and therefore the contribution to the Wiener index of T^* of all such vertex pairs is $W(T)$.

Case 2: $x \in V(T)$, e.g. $x = v_j$, and $y \in V(T^*) \setminus V(T)$ with y attached to a vertex v_i. Then we obtain for all n_i pairs $\{x,y\}$ that $d_{T^*}(x,y) = d_T(x,v_i)+1$.

Case 3: x, $y \in V(T^*) \setminus V(T)$ with x attached to a vertex v_i and y attached to a vertex v_j, $i \neq j$. Then $d_{T^*}(x,y) = d_T(v_i,v_j)+2$, and there are exactly $n_i n_j$ such pairs $\{x,y\}$.

Case 4: x, $y \in V(T^*) \setminus V(T)$ with both attached to a vertex v_i. Then $d_{T^*}(x,y) = 2$ and the number of such pairs is $\binom{n_i}{2}$.

Altogether we get

$$\begin{aligned} W(T^*) &= W(T) + \sum_{i=1}^{n}\sum_{j=1}^{n} n_i(d_T(v_i,v_j)+1) + \sum_{1\le i<j\le n} n_i n_j (d_T(v_i,v_j)+2) + \sum_{i=1}^{n} 2\binom{n_i}{2} \\ &= W(T) + \sum_{1\le i<j\le n} (n_i+n_j)d_T(v_i,v_j) + \sum_{i=1}^{n}\sum_{j=1}^{n} n_i + \sum_{1\le i<j\le n} n_i n_j d_T(v_i,v_j) \\ &\quad + \left(\sum_{i=1}^{n} n_i\right)^2 - \sum_{i=1}^{n} n_i^2 + \sum_{i=1}^{n} n_i(n_i-1) \\ &= W(T) + \sum_{1\le i<j\le n} (n_i+n_j)d_T(v_i,v_j) + \sum_{1\le i<j\le n} n_i n_j d_T(v_i,v_j) \\ &\quad + \left(\sum_{i=1}^{n} n_i\right)^2 + (n-1)\sum_{i=1}^{n} n_i. \end{aligned}$$

Since $W(T) = \sum\limits_{1\le i<j\le n} d_T(v_i,v_j)$, we can also write the Wiener index of T^* as

$$W(T^*) = \sum_{1\le i<j\le n} (n_i+1)(n_j+1)d_T(v_i,v_j) + \left(\sum_{i=1}^{n} n_i\right)^2 + (n-1)\sum_{i=1}^{n} n_i,$$

which completes the proof. □

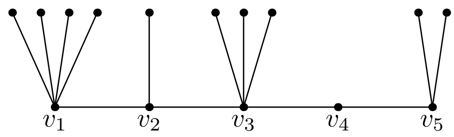


Figure 2.11

Example 2.45. Let us consider the thorn tree T^* shown in Figure 2.11. To use Theorem 2.44 for computing the Wiener index of T^* we choose the path $(v_1, v_2, v_3, v_4, v_5)$ to be the tree T and therefore $n_1 = 4$, $n_2 = 1$, $n_3 = 3$, $n_4 = 0$ and $n_5 = 2$. Thus we obtain

$$\begin{aligned} W(T^*) &= \sum_{i=1}^{5} \sum_{j=i+1}^{5} (n_i + 1)(n_j + 1)|j - i| + \left(\sum_{i=1}^{5} n_i\right)^2 + 4\sum_{i=1}^{5} n_i \\ &= 186 + 100 + 40 = 326. \end{aligned}$$

Remark 2.46. It is easily seen that in Theorem 2.44 the graph T need not be a tree, but only a finite, connected and simple graph.

Now let us consider some special cases of Theorem 2.44:

Corollary 2.47. *Let T be a tree on n vertices $v_1, v_2, \ldots, v_n$, and T^* its thorn tree with $n_i = k$, $i = 1, 2, \ldots, n$. Then*

$$W(T^*) = (k + 1)^2 W(T) + nk(nk + n - 1).$$

Proof. Simple substituting leads to

$$\begin{aligned} W(T^*) &= W(T) + \sum_{1 \le i < j \le n} 2k d_T(v_i, v_j) + k^2 \sum_{1 \le i < j \le n} d_T(v_i, v_j) \\ &\quad + \left(\sum_{i=1}^{n} k\right)^2 + (n - 1)\sum_{i=1}^{n} k \\ &= (1 + 2k + k^2)W(T) + k^2 n^2 + kn(n - 1) \\ &= (k + 1)^2 W(T) + nk(nk + n - 1). \end{aligned}$$

□

Example 2.48. Let T^* arise from the tree T of Figure 2.11 by connecting 3 vertices to each vertex of T. According to Example 2.45 the Wiener index of T is 326. Therefore we easily get

$$W(T^*) = 4 \cdot 326 + 15 \cdot 3(15 \cdot 3 + 15 - 1) = 3959.$$

To obtain two further corollaries we first have to prove the following lemma:

Lemma 2.49. *Let T be a tree with vertex set $V(T) = \{v_1, v_2, \dots, v_n\}$. Then*

$$\sum_{1\le i<j\le n} (\deg_T(v_i) + \deg_T(v_j))d_T(v_i, v_j) = 4W(T) - n(n-1) \tag{2.6}$$

and

$$\sum_{1\le i<j\le n} \deg_T(v_i)\deg_T(v_j)d_T(v_i, v_j) = 4W(T) - (n-1)(2n-1). \tag{2.7}$$

Proof. To prove the two equations we use equation (2.1) in a more generalized form, which means that every pair $\{v_i, v_j\}$ is associated with a weight ω_{ij}. Since equation (2.1) can be rewritten as

$$W(T) = \sum_{e=(v_a,v_b)\in E(T)} \sum_{v_i\in V(T_a)} \sum_{v_j\in V(T_b)} 1$$

using the notation defined in Theorem 2.6, we obtain by similar considerations

$$W_\omega(T) := \sum_{1\le i<j\le n} \omega_{ij} d_T(v_i, v_j) = \sum_{e=(v_a,v_b)\in E(T)} \sum_{v_i\in V(T_a)} \sum_{v_j\in V(T_b)} \omega_{ij}.$$

Let $\omega_{ij} = \deg_T(v_i) + \deg_T(v_j)$. Then

$$\begin{aligned}
&\sum_{1\le i<j\le n} (\deg_T(v_i) + \deg_T(v_j))d_T(v_i, v_j) = \\
&\qquad = \sum_{e=(v_a,v_b)\in E(T)} \left(\sum_{v_i\in V(T_a)} \sum_{v_j\in V(T_b)} \deg_T(v_j) + \sum_{v_i\in V(T_a)} \sum_{v_j\in V(T_b)} \deg_T(v_j) \right) \\
&\qquad = \sum_{e=(v_a,v_b)\in E(T)} \left(n_{v_b}(e) \sum_{v_i\in V(T_a)} \deg_T(v_i) + n_{v_a}(e) \sum_{v_j\in V(T_b)} \deg_T(v_j) \right) \\
&\qquad = \sum_{e=(v_a,v_b)\in E(T)} (n_{v_b}(e)[2(n_{v_a}(e)-1)+1] + n_{v_a}(e)[2(n_{v_b}(e)-1)+1]) \\
&\qquad = \sum_{e=(v_a,v_b)\in E(T)} [4n_{v_a}(e)n_{v_b}(e) - (n_{v_a}(e) + n_{v_b}(e))] \\
&\qquad = 4W(T) - n(n-1).
\end{aligned}$$

The third equation holds as $\sum_{v\in V(G)} \deg_G(v) = 2|E(G)|$ for all finite graphs G, and furthermore we use the fact that $\deg_T(v) = \deg_{T_l}(v)$ for all vertices $v \in V(T_l \setminus \{v_l\})$ and $\deg_T(v_l) = \deg_{T_l}(v_l) - 1$, $l = a, b$.

Now let $\omega_{ij} = \deg_T(v_i)\deg_T(v_j)$. Then

$$\sum_{1\le i<j\le n} (\deg_T(v_i)\deg_T(v_j))d_T(v_i, v_j) = \sum_{e=(v_a,v_b)\in E(T)} \left(\sum_{v_i\in V(T_a)} \deg_T(v_i) \sum_{v_j\in V(T_b)} \deg_T(v_j) \right)$$

$$
\begin{aligned}
&= \sum_{e=(v_a,v_b)\in E(T)} [2(n_{v_a}(e)-1)+1][2(n_{v_b}(e)-1)+1] \\
&= \sum_{e=(v_a,v_b)\in E(T)} [4n_{v_a}(e)n_{v_b}(e) - 2(n_{v_a}(e)+n_{v_b}(e))+1] \\
&= 4W(T) - (n-1)(2n-1),
\end{aligned}
$$

which completes the proof. □

Corollary 2.50. *Let T be a tree of order n and T^* its thorn tree with parameters $n_i = \deg_T(v_i)$, $i = 1, 2, \ldots, n$. Then*

$$W(T^*) = 9W(T) + (n-1)(3n-5).$$

Proof. Using Theorem 2.44 and Lemma 2.49 we obtain

$$
\begin{aligned}
W(T^*) &= W(T) + \sum_{1\le i<j\le n} (\deg_T(v_i)+\deg_T(v_j))d_T(v_i,v_j) \\
&\quad + \sum_{1\le i<j\le n} \deg_T(v_i)\deg_T(v_j)d_T(v_i,v_j) + \left(\sum_{i=1}^{n} \deg_T(v_i)\right)^2 \\
&\quad + (n-1)\sum_{i=1}^{n} \deg_T(v_i) \\
&= W(T) + 4W(T) - n(n-1) + 4W(T) - (n-1)(2n-1) \\
&\quad + 4(n-1)^2 + 2(n-1)^2 \\
&= 9W(T) + (n-1)(3n-5).
\end{aligned}
$$

□

Corollary 2.51. *Let T be a tree on n vertices and T^* its thorn tree with parameters $n_i = m - \deg_T(v_i) \ge 0$, $i = 1, 2, \ldots, n$. Then*

$$W(T^*) = (m-1)^2W(T) + [(m-1)n+1]^2.$$

Proof. In the same manner as in Corollary 2.50 we compute

$$
\begin{aligned}
W(T^*) &= W(T) + \sum_{1\le i<j\le n} (2m - \deg_T(v_i) - \deg_T(v_j))d_T(v_i,v_j) \\
&\quad + \sum_{1\le i<j\le n} (m-\deg_T(v_i))(m-\deg_T(v_j))d_T(v_i,v_j) \\
&\quad + \left(\sum_{i=1}^{n}(m-\deg_T(v_i))\right)^2 + (n-1)\sum_{i=1}^{n}(m-\deg_T(v_i)) \\
&= W(T) + 2mW(T) - (4W(T) - n(n-1)) + m^2W(T)
\end{aligned}
$$

$$
\begin{aligned}
&- m(4W(T) - n(n-1)) + 4W(T) - (n-1)(2n-1) \\
&+ (mn - 2(n-1))^2 + (n-1)(mn - 2(n-1)) \\
= {}& (m-1)^2 W(T) + (mn)^2 - 2mn(n-1) + (n-1)^2 \\
= {}& (m-1)^2 W(T) + [(m-1)n + 1]^2.
\end{aligned}
$$

□

Another special case of thorn trees are the so-called caterpillars. In [2] a formula for regular caterpillars is presented.

Definition 2.8. If T is a path, its thorn tree T^* is called a *caterpillar.*

We denote a caterpillar by $T^*(a,b)$ if all vertices which are not leaves have the same degree $a > 1$ and $b > 0$ is the number of non-leaves.

Corollary 2.52. *Let $T^*(a,b)$ be as before. Then*

$$W(T^*(a,b)) = \frac{(a-1)b}{6}[(a-1)(b-1)(b+7) + 6(a+1)] + 1.$$

Proof. Substituting into the formula of Corollary 2.51 we get

$$
\begin{aligned}
W(T^*(a,b)) &= (a-1)^2 W(P_b) + [(a-1)b + 1]^2 \\
&= (a-1)^2 \binom{b+1}{3} + (a-1)^2 b^2 + 2(a-1)b + 1 \\
&= (a-1)b \left[(a-1) \left(\frac{b^2-1}{6} + b \right) + 2 \right] + 1 \\
&= \frac{(a-1)b}{6}[(a-1)(b^2 + 6b - 1) + 12] + 1 \\
&= \frac{(a-1)b}{6}[(a-1)(b-1)(b+7) + 6(a+1)] + 1.
\end{aligned}
$$

□

2.2.2 A k-subdivision of a tree and its Wiener index

By increasing the length of the segments in a tree we obtain another concept of growing trees (see [5]):

Definition 2.9. Let T be a tree. T' is called *k-subdivision of T* if T' arises from T by replacing every edge of T by a path of length $k+1$.

Notice that the order of T' is $n' = k(n-1) + n$ and the degree of each new vertex is exactly 2, whereas $\deg_{T'}(v) = \deg_T(v)$ for all $v \in V(T)$.

Theorem 2.53. *Let T be a tree on n vertices and T' its k-subdivision. Then*

$$W(T') = (k+1)^3 W(T) + \binom{n'+1}{3} - (k+1)^3 \binom{n+1}{3}$$

with n' the order of T'.

Proof. Applying formula 2.2 to T' we get

$$\begin{aligned} W(T') &= \binom{n'+1}{3} - \sum_{v \in V(T')} \sum_{1 \le i < j < l \le \deg_{T'}(v)} |V(T'_i)| \cdot |V(T'_j)| \cdot |V(T'_l)| \\ &= \binom{n'+1}{3} - \sum_{v \in V(T)} \sum_{1 \le i < j < l \le \deg_{T}(v)} (k+1)^3 |V(T_i)| \cdot |V(T_j)| \cdot |V(T_l)| \\ &= \binom{n'+1}{3} + (k+1)^3 W(T) - (k+1)^3 \binom{n+1}{3}. \end{aligned}$$

□

Example 2.54. Let T be the tree shown in Figure 2.1 and T' its 3-subdivision. As we already know that $W(T) = 250$ we can easily compute the Wiener index of T' by using Theorem 2.53. Thus we obtain

$$W(T') = 4^3 \cdot 250 + \binom{50}{3} - 4^3 \binom{14}{3} = 12304.$$

In [18] Polansky and Bonchev also consider a tree T_1 which is obtained from T by 1-subdividing only one edge $e = (u, v)$ of T:

Theorem 2.55. *Let T be a tree of order n and $e = (u, v) \in E(T)$. Furthermore let T_1 be the tree as described above. Then*

$$W(T_1) = W(T) + \frac{1}{2}[d_T(u) + d_T(v) + n_u(e) + 2n_u(e)n_v(e) + n_v(e)].$$

Proof. Let the new vertex between u and v be denoted by w. Considering two vertices x and $y \in V(T_1)$ we distinguish between three cases:
If $x, y \in V(T_l)$, $l = u, v$, we have $d_{T_1}(x, y) = d_T(x, y)$.
If $x \in V(T_u)$ and $y \in V(T_v)$, we obtain $d_{T_1}(x, y) = d_T(x, y) + 1$.
If $x \in V(T_l)$, $l = u, v$, and $y = w$, we get $d_{T_1}(x, y) = d_T(x, l) + 1$.
Furthermore it is easy to see that

$$d_T(u) = d_{T_u}(u) + \sum_{x \in V(T_v)} d_T(u, v) = d_{T_u}(u) + d_{T_v}(v) + n_v(e)$$

and analogously

$$d_T(v) = d_{T_u}(u) + d_{T_v}(v) + n_u(e).$$

Altogether we obtain

$$\begin{aligned}
W(T_1) = & \sum_{\{x,y\}\subseteq V(T_u)} d_T(x,y) + \sum_{\{x,y\}\subseteq V(T_v)} d_T(x,y) + \sum_{x\in V(T_u)} \sum_{y\in V(T_v)} (d_T(x,y)+1) \\
& + \sum_{x\in V(T_u)} (d_T(x,u)+1) + \sum_{x\in V(T_v)} (d_T(x,v)+1) \\
= & W(T) + n_u(e)n_v(e) + d_{T_u}(u) + n_u(e) + d_{T_v}(v) + n_v(e) \\
= & W(T) + n_u(e)n_v(e) + \frac{1}{2}[d_T(u) + d_T(v) - n_u(e) - n_v(e)] + n_u(e) + n_v(e) \\
= & W(T) + \frac{1}{2}[d_T(u) + d_T(v) + n_u(e) + 2n_u(e)n_v(e) + n_v(e)].
\end{aligned}$$

□

Example 2.56. We again choose T to be the tree in Figure 2.1. Furthermore let T_1 be the tree which is obtained by 1-subdividing $e = (v_4, v_5)$. Since $d_T(v_4) = 27$, $d_T(v_5) = 30$, $n_{v_4}(e) = 8$ and $n_{v_5}(e) = 5$, we get

$$W(T_1) = 250 + \frac{1}{2}(17 + 30 + 8 + 2 \cdot 8 \cdot 5 + 5) = 325$$

according to Theorem 2.55.

Chapter 3

Lower and upper bounds

Since calculating the Wiener index of a graph can be computationally expensive, it is of some interest to know the extreme values of the Wiener index, particularly of graphs belonging to certain classes.

3.1 Bounds for general graphs

Some very basic bounds for the Wiener index are given in [11]. The first inequality mentioned here shows how the Wiener index of a graph and of its subgraph are related.

Theorem 3.1. *Let G be a connected graph and $e \in E(G)$. Furthermore let G' be the graph with vertex set $V(G)$ and edge set $E(G) \setminus \{e\}$. Then*

$$W(G) < W(G').$$

Proof. Let $e = (u, v)$. Then for each vertex pair x, y with e lying on the shortest path between x and y we obviously have $d_G(x, y) \leq d_{G'}(x, y)$. Since $d_G(u, v) < d_{G'}(u, v)$, we obtain $W(G) < W(G')$. □

This immediately leads to the following theorem about graphs and their spanning trees:

Theorem 3.2. *Let G be a connected graph and T its spanning tree. Then*

$$W(G) \leq W(T)$$

with equality if and only if G is a tree.

A lower bound for the Wiener index of an arbitrary graph is given by the next theorem:

Theorem 3.3. *Let G be a connected graph. Then*

$$W(G) \geq \frac{n(n-1)}{2}.$$

Proof. The Wiener index of the complete graph K_n on n vertices can be easily computed as

$$W(K_n) = \sum_{i=1}^{n-1} i = \frac{n(n-1)}{2}.$$

According to Theorem 3.1 each subgraph G of K_n with $E(G) \subset E(K_n)$ has Wiener index less than the Wiener index of K_n. Since each graph of order n is a subgraph of the complete graph, the inequality holds. □

In [8] closer bounds are given if the number of edges is fixed.

Theorem 3.4. *Let G be a connected graph with n vertices and m edges. Then*

$$n(n-1) - m \leq W(G) \leq \frac{n^3 + 5n - 6}{6} - m.$$

Proof. In order to prove the first inequality we consider the distance between two vertices u and v. If u and v are neighbours, we obtain $d_G(u,v) = 1$ and otherwise $d_G(u,v) \geq 2$. Since $|E(G)| = m$, there are exactly m unordered vertex pairs with distance 1 and $\binom{n}{2} - m$ unordered vertex pairs with distance greater than 1. Thus we obtain

$$W(G) = \sum_{\{u,v\} \subseteq V(G)} d_G(u,v) \geq m + 2\left[\binom{n}{2} - m\right] = n(n-1) - m.$$

The second inequality is shown by induction on n and $n - 1 \leq m \leq \binom{n}{2}$. For $n = 2$ the inequality holds since K_2 is the only connected graph and $W(K_2) = 1 \leq \frac{8+10-6}{6} - 1 = 1$. Now let us assume that the second inequality holds for all connected graphs of order n. Let T be a tree on $n+1$ vertices and v a leaf of T. Furthermore let T' be the subtree of T induced by $V(T) \setminus \{v\}$. Then the inequality holds for T' and we obtain

$$\begin{aligned} W(T) = W(T') + d_T(v) &\leq \frac{n^3 + 5n - 6}{6} - (n-1) + \sum_{i=1}^{n} i \\ &= \frac{(n+1)^3 + 5(n+1) - 6}{6} - n. \end{aligned}$$

Let us assume that the inequality holds for all connected graphs with n vertices and $m \geq n-1$ edges. We consider the connected graph G with n vertices and $m+1$ edges. Since G is no tree, it contains an edge e such that G' with vertex set $V(G)$ and edge set $E(G) \setminus \{e\}$ is a connected subgraph of G. According to Theorem 3.1 we obtain $W(G) \leq W(G') - 1$. As the assumption holds for G' we have

$$W(G) \leq W(G') - 1 \leq \frac{n^3 + 5n - 6}{6} - m - 1,$$

which completes the proof. □

Corollary 3.5. *Among all connected graphs of order n the path P_n maximizes the Wiener index.*

Proof. Let G be a connected graph with n vertices and m edges. Since G is connected, we have $n-1 \leq m$. Due to Theorem 3.4 we have

$$\begin{aligned} W(G) &\leq \frac{n^3+5n-6}{6} - m \leq \frac{n^3+5n-6}{6} - (n-1) \\ &= \frac{n^3-n}{6} = \binom{n+1}{3} = W(P_n). \end{aligned}$$

□

3.2 Bounds for trees

Because of Corollary 3.5 we already have an upper bound for the Wiener index of trees. By Corollary 3.6 we also obtain a lower bound.

Corollary 3.6. *Among all trees of order n the star S_n minimizes the Wiener index.*

Proof. Since the number of edges in a tree of order n is $m = n-1$, we obtain by Theorem 3.4

$$(n-1)^2 \leq W(T).$$

Thus, as $W(S_n) = (n-1)^2$, the star S_n has the minimal Wiener index among all trees of order n. □

Another question that immediately arises is: Which extremal values of the Wiener index can be obtained if the order of the tree and its maximum degree or even its entire degree sequence is given? In the following two subsections we will take a closer look at these problems.

3.2.1 Trees with given maximum degree

As the maximum degree in P_n is 2, it is obvious that P_n still maximizes the Wiener index among all trees of order n and maximum degree at most $\Delta \geq 2$. Therefore this subsection will be mainly devoted to solve the following problem:

Problem 3.1. What trees minimize the Wiener index among all trees of given order n and maximum degree at most $\Delta \geq 2$?

Remark 3.7. If $\Delta = 2$, the only tree with degree at most 2 and therefore the optimal solution is the path P_n. Also if $n \leq \Delta+1$, the problem gets trivial and the optimal solution is the star S_{n+1}. Thus Problem 3.1 is only of interest if $\Delta \geq 3$ and $n > \Delta + 1$.

In order to discuss Problem 3.1 we have to consider a special class of trees described in the definition below (see [9]).

Definition 3.2. Let $\Delta \geq 3$ and $R \in \{\Delta - 1, \Delta\}$. Then $\mathscr{T}(R, \Delta)$ is defined as the set of all trees T which can be embedded in the plane as follows:

Let $n = |V(T)|$, $M_0(R, \Delta) = 1$ and $M_i(R, \Delta) = 1 + R\sum_{j=0}^{k-1}(\Delta - 1)^j$ for all $i \geq 1$. Furthermore let

$$M_k(R, \Delta) \leq n < M_{k+1}(R, \Delta)$$

for some $k \geq 0$ and

$$n - M_k(R, \Delta) = m(\Delta - 1) + r$$

with $m \geq 0$ and $0 \leq r < \Delta - 1$. Then

1. all vertices of T lie on some line $\mathbb{R} \times \{i\}$ with $0 \leq i \leq k + 1$,
2. on line $\mathbb{R} \times \{0\}$ there is just one single vertex which has exactly $\min\{n - 1, R\}$ neighbours lying all on line $\mathbb{R} \times \{1\}$,
3. on line $\mathbb{R} \times \{i\}$, $1 \leq i \leq k-1$, every vertex has a unique neighbour on line $\mathbb{R} \times \{i-1\}$ and $\Delta - 1$ neighbours on line $\mathbb{R} \times \{i + 1\}$,
4. if $v_1, \ldots, v_{m+1}$ are the $m + 1$ leftmost vertices on line $\mathbb{R} \times \{k\}$ such that v_a lies left of v_b for $a < b$, then v_j has $\Delta - 1$ neighbours on line $\mathbb{R} \times \{k + 1\}$, $1 \leq j \leq m$, and v_{m+1} has r neighbours on line $\mathbb{R} \times \{k + 1\}$.

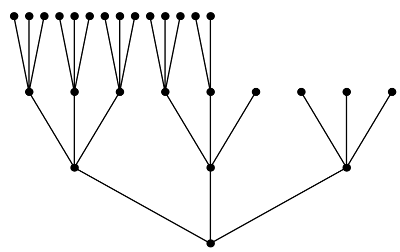


Figure 3.1: The tree $T \in \mathscr{T}(3, 3)$ with $n = 27$.

Example 3.8. An example of an element of $\mathscr{T}(R, \Delta)$ with $R = \Delta = 3$ is shown in Figure 3.1.

Remark 3.9. For every $n \in \mathbb{N}$ the set $\mathscr{T}(R, \Delta)$ contains a unique tree T of order n up to isomorphism.

Notice that $M_i(R, \Delta)$ is the number of vertices being assigned to the lines 0 up to i.

Definition 3.3. Let T be a tree on n vertices and $v \in V(T)$. A maximal subtree B containing v as a leaf is called a *branch of T at v*. The *weight* $bw(B)$ of B is defined as the number of edges in B and the *branch-weight* $bw(v)$ of v is the maximum of the weights of all branches at v. Then it is easy to see that $n = \sum_{i=1}^{k} bw(B_i)$ with B_i, $i = 1, \ldots, k$, being the k branches of T at v.

Furthermore the *centroid* $C(T)$ of T is the set of vertices in T with minimum branch-weight.

In [5] the following characterisation of centroids is mentioned:

Theorem 3.10. *Let T be a tree of order n and $C(T)$ its centroid. Then one of the following holds:*

(i) $C(T) = \{c\}$ and $bw(c) \leq \frac{n-1}{2}$,

(ii) $C(T) = \{c_1, c_2\}$ and $bw(c_1) = bw(c_2) = \frac{n}{2}$.

Proof. Assume that $|C(T)| \geq 3$, then there exist pairwise distinct c_1, c_2, $c_3 \in C(T)$. It is obvious that $bw(c_1) = bw(c_2) = bw(c_3)$. Let B_1 be the branch of c_1 with $bw(B_1) = bw(c_1)$. Then we distinguish the following cases:

(i) $c_2, c_3 \notin V(B_1)$,

(ii) $c_2 \notin V(B_1)$ and $c_3 \in V(B_1)$ (the case $c_2 \in V(B_1)$ and $c_3 \notin V(B_1)$ is the same due to symmetry),

(iii) $c_2, c_3 \in V(B_1)$.

For cases (i) and (ii) we immediately obtain that there exists a branch B_2 of c_2 with $V(B_1) \cup \{c_1\} \subseteq V(B_2)$. Thus we have $bw(B_2) > bw(B_1) = bw(c_1) = bw(c_2)$ which is a contradiction.

Now let $c_2, c_3 \in V(B_1)$ (case (iii)). If there exists another branch B_1' of c_1 with $bw(B_1') = bw(B_1)$, we obviously have $c_2, c_3 \notin V(B_1')$ and therefore we can use the argumentation of cases (i) and (ii) to obtain a contradiction. Thus for all branches $B_1' \neq B_1$ of c_1 we have $bw(B_1') < bw(B_1)$. Since c_2 and c_3 are in the same branch of c_1, there exists a vertex v which is a neighbour of c_1 and lies on the paths from c_1 to c_2 and from c_1 to c_3. That indicates that $v \in V(B_1)$ and either $v \neq c_2$ or $v \neq c_3$. W.l.o.g. $v \neq c_2$. Since there exists a path $c_1 \to v \to c_2$, we obtain that there also exists a branch B_2 of c_2 with $c_1, v \in V(B_2)$. Let B_v be a branch of v. If $c_1 \in V(B_v)$, we get $V(B_v) \subset V(B_2)$ and thus $bw(B_v) < bw(B_2) \leq bw(c_2) = bw(c_1)$. If $c_1 \notin V(B_v)$, we have $V(B_v) \subset V(B_1)$ and therefore $bw(B_v) < bw(B_1) = bw(c_1)$. Hence we obtain $bw(v) < bw(c_1)$, which is a contradiction to the assumption that $c_1 \in C(T)$.

Therefore $|C(T)| \leq 2$. Let $C(T) = \{c\}$ and furthermore we assume $bw(c) > \frac{n-1}{2}$. Let B_1, B_2, …, B_k be the branches at c and w.l.o.g. let $bw(B_1) = bw(c)$. Furthermore we denote the vertex of B_1 connected with c by u. As $E(T) = n - 1$, we get that

$$\sum_{i=2}^{k} bw(B_i) < \frac{n-1}{2}.$$

Considering the branch B_a of u containing B_2, ..., B_k and c we obtain $bw(B_a) < \frac{n-1}{2} + 1$ and therefore

$$bw(B_a) \leq bw(c).$$

On the other hand, as the other branches of u are subtrees of B_1 without c, their branch-weight is smaller than $bw(B_1)$. Altogether we get

$$bw(u) \leq bw(c),$$

which is a contradiction to $C(T) = \{c\}$.

In case $C(T) = \{c_1, c_2\}$ it is easy to see that for c_i the branch B_i with maximum weight has to contain c_j with $i \neq j$. Furthermore there are no further vertices on the path between c_1 and c_2 (as otherwise such a vertex would lie in both B_1 and B_2 and therefore the weight of such a vertex would be less than $bw(c_1)$). Since all edges except (c_1, c_2), which is counted twice, are counted exactly once, we get

$$bw(c_1) + bw(c_2) = n - 1 + 1 = n$$

and thus $bw(c_1) = bw(c_2) = \frac{n}{2}$, which completes the proof. □

In order to find a solution to Problem 3.1 we have to prove the following lemma, which can be found in [9], first.

Lemma 3.11. *Let $T \in \mathscr{T}(R, \Delta)$ be of order n with $M_k(R, \Delta) < n < M_{k+1}(R, \Delta)$ for some $k \geq 1$. Furthermore let $T_0 \in \mathscr{T}(R, \Delta)$ be of order $M_k(R, \Delta)$ and T arise from T_0 by attaching $n - M_k(R, \Delta)$ new vertices which lie on the line $\mathbb{R} \times \{k+1\}$ to the vertices of T_0 lying on the line $\mathbb{R} \times \{k\}$. Then either $W(T) < W(T')$ or T and T' are isomorphic.*

Proof. Let T' be the tree with minimum Wiener index among all trees which arise from T_0 in the manner described above. Then we show that T' and T are isomorphic.

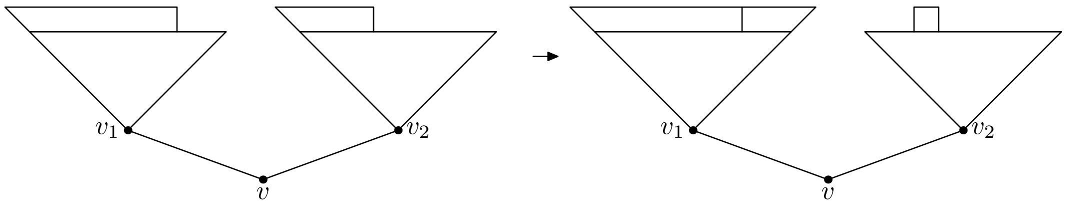
Let $v \in V(T_0)$ lie on the line $\mathbb{R} \times \{i\}$ with $0 \leq i \leq k$. Furthermore the subtree of T' containing v and only vertices lying on a line $\mathbb{R} \times \{j\}$ with $j > i$ is denoted by T_v. We call T_v full (empty) if all vertices of T_v on line $\mathbb{R} \times \{k\}$ have degree Δ (1).

We choose the following planar embedding of T': For each v lying on line $\mathbb{R} \times \{i\}$, $i \leq k - 1$, beginning with $i = k - 1$ we consider the subtrees of all neighbours of v on line $\mathbb{R} \times \{i + 1\}$. We arrange the subtrees according to the number of their vertices in a decreasing manner, the subtree with most vertices lying leftmost.

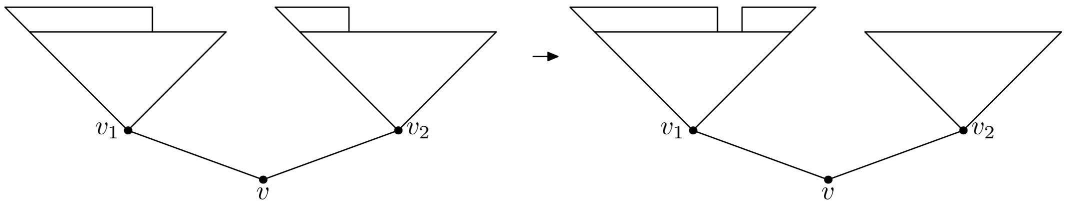
Claim: Let the vertex v lie on line $\mathbb{R} \times \{i\}$, $i \leq k - 1$. Furthermore we denote the neighbours of v on line $\mathbb{R} \times \{i+1\}$ by v_1, ..., v_l. Then there exists at most one tree among T_{v_1}, ..., T_{v_l} which is neither full nor empty.

If the claim holds, with this particular planar embedding of T' it is easy to see that T' and T are isomorphic.

Proof of the claim: In order to prove the claim by contradiction we assume that there exists a vertex v lying on line $\mathbb{R} \times \{i\}$ with maximum i, $i \leq k - 1$, such that v has at least two neighbours v_1 and v_2 on line $\mathbb{R} \times \{i + 1\}$ with T_{v_1}, T_{v_2} being neither full nor



(a) Case $p = (\Delta - 1)^{k-i} - |V_1|$.



(b) Case $p = |V_2|$.

Figure 3.2: How T'' arises from T'.

empty. W.l.o.g. we assume that $|V(T_{v_1})| \geq |V(T_{v_2})|$, which also means that v_1 lies left of v_2 according to the embedding described above. Since the claim holds for all $j > i$, there is at most one vertex in $V(T_a)$, $a = 1, 2$, on line $\mathbb{R} \times \{k\}$ of degree $\neq 1$, Δ.

Furthermore we define

$$V_1 = \{x \in T_{v_1} \,:\, x \text{ lies on line } \mathbb{R} \times \{k+1\}\}$$

and V_2 analogously. Thus we get

$$1 \leq |V_1| \leq |V_2| < (\Delta - 1)^{k-i}.$$

Let $p = \min\{(\Delta - 1)^{k-i} - |V_1|, |V_2|\}$, then obviously $1 \leq p$. The idea is to take p vertices of V_2 and to rearrange them such that they are in $V_1 \setminus V_2$. Therefore let $x_1, \ldots, x_p$ be the leftmost vertices of V_2, and furthermore with $q = \lceil \frac{p}{\Delta-1} \rceil$ let $y_1, \ldots, y_q$ be the rightmost vertices in $V(T_{v_1})$ on line $\mathbb{R} \times \{k\}$ and $z_1, \ldots, z_q$ be the leftmost vertices in $V(T_{v_2})$ on the same line, all counted from left to right. Then we define the tree T'' with vertex set $V(T')$ and edge set

$$\big(E(T') \setminus \{(x_j, z_{\lceil \frac{j}{\Delta-1} \rceil}) \,:\, 1 \leq j \leq p\}\big) \cup \{(x_j, y_{\lceil \frac{j+b}{\Delta-1} \rceil}) \,:\, 1 \leq j \leq p\}$$

with $b = \deg_{T'}(y_1) - 1$ (see Figure 3.2).

If $x, y \in V(T') \setminus \{x_1, \ldots, x_p\}$, the distance between x and y is the same in T'' as in T'. Moreover the sum of the distances between all vertex pairs $x \in \{x_1, \ldots, x_p\}$ and $y \in (V(T') \setminus (V_1 \cup V_2)) \cup \{x_1, \ldots, x_p\}$ is also the same in T'' as in T'. Thus we get

$$W(T') - W(T'') = \sum_{\substack{x \in \{x_1, \ldots, x_p\} \\ y \in (V_1 \cup V_2) \setminus \{x_1, \ldots, x_p\}}} (d_{T'}(x, y) - d_{T''}(x, y)).$$

Considering the subtree of T'' induced by $V(T_{v_1}) \cup \{x_1, \ldots, x_p\}$ it is easy to see that it contains a tree $T^*_{v_2} \cong T_{v_2}$ such that $\{x_1, \ldots, x_p\} \subseteq V(T^*_{v_2})$. Let $V_1^* = V_1 \setminus V(T^*_{v_2})$, then

$$\sum_{\substack{x \in \{x_1,\ldots,x_p\} \\ y \in (V_1 \setminus V_1^*) \cup V_2 \setminus \{x_1,\ldots,x_p\}}} (d_{T'}(x,y) - d_{T''}(x,y)) = 0.$$

Thus we obtain

$$\begin{aligned} W(T') - W(T'') &= \sum_{\substack{x \in \{x_1,\ldots,x_p\} \\ y \in V_1^*}} (d_{T'}(x,y) - d_{T''}(x,y)) \\ &\geq \sum_{\substack{x \in \{x_1,\ldots,x_p\} \\ y \in V_1^*}} (2(k+1-i) - 2(k-i)) > 0, \end{aligned}$$

which is a contradiction to the assumption concerning the Wiener index of T' and therefore the claim holds. □

In Lemma 3.12 we use the following notation: Let $v \in V(T)$ with T a tree. If the centroid $C(T) \neq \{v\}$, then let T_v be the subtree of T induced by

$$\{u \in V(T) \,:\, d_T(u,c) = d_T(u,v) + d_T(v,c) \text{ for all } c \in C(T)\}$$

(compare definition of collinear). If $C(T) = \{v\}$, then $T_v = T$.

Lemma 3.12. *Let T be a tree on n vertices and maximum degree at most Δ, $\Delta \geq 3$, such that $W(T) \leq W(T')$ for all trees T' on n vertices and maximum degree at most Δ. Furthermore let $v \in V(T)$. Then $T_v \in \mathscr{T}(\Delta-1,\Delta)$ if $C(T) \neq \{v\}$, and $T_v \in \mathscr{T}(\Delta,\Delta)$ if $C(T) = \{v\}$.*

Proof. If $n \leq \Delta + 1$, the tree with minimum Wiener index is the star S_n, as mentioned in Remark 3.7, and hence the lemma holds in this case.

Thus let $n > \Delta + 1$. In order to prove the lemma by induction, we define $h(T_v)$ as the maximum distance of v to a leaf of T_v. In case $h(T_v) = 0$, we have $V(T_v) = \{v\}$ and thus $T_v \in \mathscr{T}(\Delta,\Delta)$. If $h(T_v) = 1$, T_v is a star with $\deg_{T_v}(v) \leq \Delta - 1$ and, as a result, the lemma holds too.

Now let $h(T_v) \geq 2$.

Claim: Let $(x_1, x_2, \ldots, x_l)$ be a path in T_v such that $x_1 = v \in C(T)$. Then we get $\deg_T(x_j) = \Delta$ for all $1 \leq j \leq l-2$.

If the claim holds, we distinguish two different cases:

- Case 1: There exist two leaves w_1 and w_2 in T_v such that $d_T(v,w_2) \geq d_T(v,w_1) + 2$.

- Case 2: For all leaves w_1 and w_2 in T_v their distance to v differs by at most 1.

Ad case 1: We assume that w_1 has minimum distance d_1 from v among all leaves of T_v and $d_2 := d_T(w_2, v) = h(T_v)$. As for all $x \neq v$ the subtree T_x of T_v is in $\mathscr{T}(R, \Delta)$ with $R \in \{\Delta - 1, \Delta\}$ by induction, the distance between x and a leaf of T_x differs by at most 1. Therefore the only subtree containing w_1 and w_2 is T_v itself. Let v_1 and v_2, $v_1 \neq v_2$, be two neighbours of v in T_v such that v_i lies on the path between v and w_i, $i = 1, 2$. By induction, $T_{v_i} \in \mathscr{T}(\Delta - 1, \Delta)$ for $i = 1, 2$. Furthermore let u be the neighbour of v_2 which lies on the path between v_2 and w_2. Therefore we obtain that in the planar embedding of T_{v_1} and T_{v_2} according to Definition 3.2 v_1 and v_2 lie on line $\mathbb{R} \times \{0\}$, u lies on line $\mathbb{R} \times \{1\}$, w_i lies on line $\mathbb{R} \times \{d_i - 1\}$ for $i = 1, 2$ and all other vertices in $V(T_{v_1})$ and $V(T_{v_2})$ lie on the lines $\mathbb{R} \times \{j\}$ with $1 \leq j \leq d_1$ and $1 \leq j \leq d_2 - 1$, respectively. W.l.o.g. let w_2 be the vertex lying leftmost on line $\mathbb{R} \times \{d_2 - 1\}$ in T_{v_2}. Furthermore let V_1 be the set of vertices on line $\mathbb{R} \times \{d_1\}$ in T_{v_1}, let V_2 be the set of vertices on line $\mathbb{R} \times \{d_2 - 1\}$ in T_{v_2} and let $V_3 = V(T_{v_2}) \setminus V(T_u)$.

Then we easily obtain

$$|V(T_u)| + |V_3| = |V(T_{v_2})| \leq bw(v) \leq bw(v_2) = |V(T) \setminus V(T_{v_2})|.$$

Hence if $|V(T_{v_1})| < |V(T_u)|$, we get

$$|V_3| < |V(T) \setminus V(T_{v_2})| - |V(T_{v_1})| = |V(T) \setminus (V(T_{v_2}) \uplus V(T_{v_1}))|.$$

Now we consider the tree T' with vertex set $V(T)$ and edge set

$$(E(T) \setminus \{(v, v_1), (v_2, u)\}) \cup \{(v, u), (v_2, v_1)\}.$$

Then

$$W(T) - W(T') = (|V(T_u)| - |V(T_{v_1})|)\ (|V(T) \setminus (V(T_{v_1}) \cup V(T_{v_2}))| - |V_3|) > 0,$$

which is a contradiction.

Therefore $|V(T_{v_1})| \geq |V(T_u)|$. Thus we obtain

$$\sum_{i=0}^{d_2-3} (\Delta - 1)^i < |V(T_u)| \leq |V(T_{v_1})| < \sum_{i=0}^{d_1} (\Delta - 1)^i,$$

which implies $d_1 > d_2 - 3$. According to the definition of d_1 and d_2, we get $d_1 = d_2 - 2$. Hence $T_{v_1} - V_1 \cong T_u - V_2$.

As w_2 is the leftmost vertex lying on line $\mathbb{R} \times \{d_1 + 1\}$ of T_{v_2}, T_u must be full if there is any vertex of V_2 lying in V_3, which would be a contradiction to $|V(T_{v_1})| \geq |V(T_u)|$ and T_{v_1} not full. Thus we obtain $V_2 \subseteq V(T_u)$, and furthermore $|V_1| \geq |V_2|$. This implies that

$$|V_3| \leq 1 + (\Delta - 2) \sum_{i=0}^{d_1 - 1} (\Delta - 1)^i.$$

In order to construct a tree T' such that $W(T) > W(T')$, analogously to the proof of Lemma 3.11 we define $p = \min\{(\Delta - 1)^{d_1} - |V_1|, |V_2|\} \geq 1$. Let $x_1, \ldots, x_p$ be the p leftmost vertices in V_2. For $q = \lceil \frac{p}{\Delta-1} \rceil$ let $y_1, \ldots, y_q$ be the rightmost q vertices on line $\mathbb{R} \times \{d_1 - 1\}$ of T_{v_1} and $z_1, \ldots, z_q$ be the leftmost q vertices on line $\mathbb{R} \times \{d_1\}$ of T_{v_2}, all counted from left to right. Then we define T' to be the tree with vertex set $V(T)$ and edge set

$$\big(E(T) \setminus \{(x_j, z_{\lceil \frac{j}{\Delta-1} \rceil}) \,:\, 1 \le j \le p\}\big) \cup \{(x_j, y_{\lceil \frac{j+b}{\Delta-1} \rceil}) \,:\, 1 \le j \le p\}$$

with $b = \deg_{T'}(y_1) - 1$. It is obvious that the sum of the distances between all pairs of vertices in $V(T) \setminus \{x_1, \ldots, x_p\}$ is the same in T and T', and analogously the sum of the distances between all vertex pairs in $\{x_1, \ldots, x_p\}$. Thus

$$W(T) - W(T') = \sum_{\substack{x \in \{x_1,\ldots,x_p\} \\ y \in V(T) \setminus \{x_1,\ldots,x_p\}}} (d_T(x,y) - d_{T'}(x,y)).$$

If $x \in \{x_1, \ldots, x_p\}$ and $y \in V(T) \setminus (V(T_{v_1}) \cup V(T_{v_2}))$, we get $d_T(x,y) - d_{T'}(x,y) = 1$. If $x \in \{x_1, \ldots, x_p\}$ and $y \in V_3$, we get $d_T(x,y) - d_{T'}(x,y) = -1$.

In case $C(T) = \{v\}$, we get $\deg_T(v) = \Delta$ by using the claim. Thus

$$|V(T) \setminus (V(T_{v_1}) \cup V(T_{v_2}))| \ge 1 + (\Delta - 2) \sum_{i=0}^{d_1-1} (\Delta - 1)^i \ge |V_3|.$$

In case $C(T) \ne \{v\}$, we also get

$$|V(T) \setminus (V(T_{v_1}) \cup V(T_{v_2}))| > |V(T) \setminus V(T_v)| \ge |V(T_v)| > |V_3|.$$

Altogether we obtain

$$\sum_{\substack{x \in \{x_1,\ldots,x_p\} \\ y \in V(T) \setminus (V(T_{v_1}) \cup V(T_u))}} (d_T(x,y) - d_{T'}(x,y)) = p(|V(T) \setminus (V(T_{v_1}) \cup V(T_{v_2}))| - |V_3|) \ge 0,$$

since $V(T_u) \cup V_3 = V(T_{v_2})$.

Considering the subtree of T' induced by the vertex set $V(T_{v_1}) \cup \{x_1, \ldots, x_p\}$ it is easy to see that it contains a subtree $T_u^* \cong T_u$ with $\{x_1, \ldots, x_p\} \subseteq V(T_u^*)$. Let $V_1^* = V_1 \setminus V(T_u^*)$, then obviously $V_1^* \ne \emptyset$. Thus we get

$$\sum_{\substack{x \in \{x_1,\ldots,x_p\} \\ y \in (V(T_{v_1}) \cup V(T_u)) \setminus V_1^*}} (d_T(x,y) - d_{T'}(x,y)) = 0$$

and finally

$$\sum_{\substack{x \in \{x_1,\ldots,x_p\} \\ y \in V_1^*}} (d_T(x,y) - d_{T'}(x,y)) \ge p|V_1^*|((2d_1 + 3) - 2d_1) > 0.$$

Therefore we get

$$W(T) - W(T') > 0,$$

which is a contradiction to the choice of T.

Ad case 2: Let $M_k(R,\Delta) \leq |V(T_v)| < M_{k+1}(R,\Delta)$ for some $k \geq 1$. Since in this case the distance from two leaves of T_v to v differs by at most 1 and the claim implies that all vertices of T except the leaves and their neighbours have degree Δ, T arises from a tree $T_0 \in \mathscr{T}(R,\Delta)$ of order $M_k(R,\Delta)$ by attaching $n - M_k(R,\Delta)$ new vertices which lie on line $\mathbb{R} \times \{k+1\}$ to the vertices of T_0 lying on line $\mathbb{R} \times \{k\}$. Using Lemma 3.11 we obtain the desired result.

Proof of the claim: Since $x_i \in V(T_v)$, $2 \leq i \leq l$, we obtain from the definition of T_v that $x_i \notin C(T)$. To prove the claim by contradiction, let $\deg_T(x_j) < \Delta$ for some $1 \leq j \leq l-2$. Furthermore we define $V_{j+2} = V(T_{x_{j+2}})$, $V_{j+1} = V(T_{x_{j+1}}) \setminus V_{j+2}$ and $V_j = V(T) \setminus (V_{j+1} \cup V_{j+2})$. Because of the definition of the centroid and T_v, it is easy to see that $|V_j| > |V_{j+1}|$. Let T'' be the tree with vertex set $V(T)$ and edge set $(E(T) \setminus \{(x_{j+1}, x_{j+2})\}) \cup \{(x_j, x_{j+2})\}$. Then we obtain

$$W(T) - W(T'') = |V_{j+2}| \, (|V_j| - |V_{j+1}|) > 0,$$

which is a contradiction to the definition of T. □

Now we can give the solution to Problem 3.1:

Theorem 3.13. *Let T be a tree of order n and maximum degree at most Δ with $\Delta \geq 3$. Then $W(T) \leq W(T')$ for all trees T' of order n and maximum degree at most Δ if and only if $T \in \mathscr{T}(\Delta,\Delta)$.*

Proof. If $C(T) = \{c\}$, it immediately follows from Lemma 3.12 that $T \in \mathscr{T}(\Delta,\Delta)$.

If $C(T) = \{c_1, c_2\}$, we know from Theorem 3.10 that $|V(T_{c_1})| = |V(T_{c_2})| = \frac{n}{2}$. Furthermore by Lemma 3.12 we obtain $T_{c_1}, T_{c_2} \in \mathscr{T}(\Delta-1,\Delta)$. Therefore T_{c_1} and T_{c_2} must be isomorphic. Because of Lemma 3.11 we get that $|V(T_{c_1})| = |V(T_{c_2})| = M_k(\Delta-1,\Delta)$ for some $k \geq 0$, which implies that $T \in \mathscr{T}(\Delta,\Delta)$.

Since the number of possible trees is finite, there has to be an optimal tree and, as we have already shown, it has to be in $\mathscr{T}(\Delta,\Delta)$. Thus if $T \in \mathscr{T}(\Delta,\Delta)$, it is the optimal tree, which completes the proof. □

The solution to Problem 3.1 leads to another question:

Problem 3.4. Which trees maximize the Wiener index among all trees of given order n (with $n \equiv 2 \mod (\Delta - 1)$) whose vertices are either leaves or of degree Δ?

A solution to this problem was also given in [9], but since it is just a special case of Problem 3.7, we will deal with it later on.

3.2.2 Trees with given degree sequence

Minimizing the Wiener index

On the following pages we consider the problem described below (see [21]).

Problem 3.5. Given an integer sequence $(d_1, \ldots, d_n)$ with

$$d_1 \geq d_2 \geq \cdots \geq d_k \geq 2 > 1 = d_{k+1} = \cdots = d_n$$

and

$$\sum_{i=1}^{n} d_i = 2(n-1),$$

we want to find a tree with degree sequence $(d_1, \ldots, d_n)$ which minimizes the Wiener index among all trees of the same degree sequence.

Definition 3.6. Let $(d_1, \ldots, d_n)$ be the degree sequence of the tree T with

$$d_1 \geq d_2 \geq \cdots \geq d_k \geq 2 > 1 = d_{k+1} = \cdots = d_n.$$

Then T is called *greedy tree* if it can be embedded in the plane as follows:

1. take the vertex v with degree d_1 as root,
2. each vertex u lies on some line i where i is the distance between the root v and u,
3. each line is filled up with vertices in decreasing degree order from left to right.

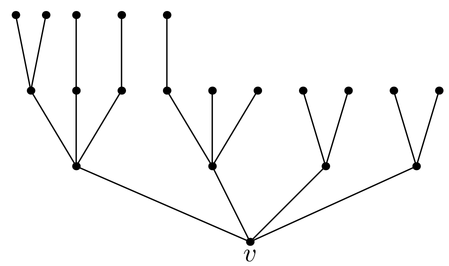


Figure 3.3: A greedy tree.

Example 3.14. Let $(4, 4, 4, 3, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ be the given degree sequence. Then its corresponding greedy tree using the plane embedding of Definition 3.6 is shown in Figure 3.3.

An equivalent definition of a greedy tree is given in the lemma below.

Lemma 3.15. *A rooted tree T with a given degree sequence is a greedy tree if the following holds:*

1. *Its root v has the largest degree.*

2. *For all leaves of T their distance to v differs by at most 1.*

3. *For any two vertices u_1 and u_2 with distance to the root $d_T(v, u_1) > d_T(v, u_2)$ follows that* $\deg_T(u_1) \leq \deg_T(u_2)$.

4. *Let u_1, u_2 be two vertices with $d_T(v, u_1) = d_T(v, u_2)$ and* $\deg_T(u_1) > \deg_T(u_2)$. *Furthermore let w_1 and w_2 be two more vertices with $d_T(v, w_i) = d_T(v, u_i) + d_T(u_i, w_i)$, $i = 1, 2$, and $d_T(v, w_1) = d_T(v, w_2)$. Then we have* $\deg_T(w_1) \geq \deg_T(w_2)$.

5. *Let u_1, u_2 be two vertices with $d_T(v, u_1) = d_T(v, u_2)$ and* $\deg_T(u_1) > \deg_T(u_2)$. *Besides let u_1', u_2' be two more vertices with $d_T(v, u_i') = d_T(v, u_i)$ and $N(u_i') \cap N(u_i) \neq \emptyset$, $i = 1, 2$; here $N(a)$ denotes the set of neighbours of the vertex a. Then we have* $\deg_T(u_1') \geq \deg_T(u_2')$.

   For all vertices u_1'', u_2'' with $d_T(v, u_1'') = d_T(v, u_2'')$ and $d_T(v, u_i'') = d_T(v, u_i') + d_T(u_i', u_i'')$, $i = 1, 2$, we also have $\deg_T(u_1'') \geq \deg_T(u_2'')$.

Before being able to show that the greedy tree is the optimal solution to Problem 3.5 we have to give some characteristics of paths and the longest path in particular of the optimal tree. In order to do so, we use the following notation:

Let T be an optimal tree according to Problem 3.5 and P be a path of T. If P is of odd length, let z be the vertex of P that separates P into two sub-paths of equal length m. Then we denote the vertices of P on the right side of z by $x_1, x_2, \ldots, x_m$ and on the left side by $y_1, y_2, \ldots, y_m$, in both cases the counting starts at z. In case P is of even length, an edge e takes the place of z and the vertices of P are labelled analogously.

Let X_i, Y_i and Z be the connected components containing x_i, y_i and z, respectively, after deleting all edges of P. Besides let $X_{>k}$ and $Y_{>k}$ denote the subtrees induced by the vertex sets $V(X_{k+1}) \cup V(X_{k+2}) \cup \ldots \cup V(X_m)$ and $V(Y_{k+1}) \cup V(Y_{k+2}) \cup \ldots \cup V(Y_m)$, respectively.

W.l.o.g. we assume that $|V(X_1)| \geq |V(Y_1)|$.

Lemma 3.16. *Let T be a tree with minimum Wiener index among all trees of the same degree sequence. Furthermore let P be a path of T with $|V(X_i)| \geq |V(Y_i)|$ for $i = 1, 2, \ldots, k$ and $k = 1, \ldots, m$. Then we can assume*

$$|V(X_{>k})| \geq |V(Y_{>k})|.$$

Proof. If $k = m$, we have $V(X_{>k}) = V(Y_{>k}) = \emptyset$ and the lemma holds. So, let $k < m$. To show the lemma by contradiction, let us assume that $|V(X_{>k})| < |V(Y_{>k})|$. Let T' be the tree which arises from T by switching $X_{>k}$ and $Y_{>k}$. It is obvious that T' and T have the same degree sequence. Furthermore we show that $W(T') \leq W(T)$. In order to do so we have to distinguish between two cases according to whether P contains a vertex z or not.

Considering the case that there is no such vertex z, the distance between two vertices in T' is as follows:

- If $x, y \in V(X_{>k}) \cup V(Y_{>k})$, we get $d_{T'}(x, y) = d_T(x, y)$.
- If $x, y \in V(T) \setminus (V(X_{>k}) \cup V(Y_{>k}))$, we have $d_{T'}(x, y) = d_T(x, y)$.
- If $x \in V(A_{>k})$ and $y \in V(A_i)$, we get $d_{T'}(x, y) = d_T(x, y) + 2i - 1$ for $i = 1, \ldots, k$ and $A = X, Y$.
- If $x \in V(X_{>k})$ and $y \in V(Y_i)$ or $x \in V(Y_{>k})$ and $y \in V(X_i)$, $i = 1, \ldots, k$, we obtain $d_{T'}(x, y) = d_T(x, y) - (2i - 1)$.

Thus

$$\begin{aligned} W(T') - W(T) &= \sum_{i=1}^{k}(2i-1)|V(X_i)||V(X_{>k})| + \sum_{i=1}^{k}(2i-1)|V(Y_i)||V(Y_{>k})| \\ &\quad - \sum_{i=1}^{k}(2i-1)|V(X_i)||V(Y_{>k})| - \sum_{i=1}^{k}(2i-1)|V(Y_i)||V(X_{>k})| \\ &= \sum_{i=1}^{k}(2i-1)(\underbrace{|V(X_i)| - |V(Y_i)|}_{\geq 0})(\underbrace{|V(X_{>k})| - |V(Y_{>k})|}_{<0}) \leq 0. \end{aligned}$$

In case there is a vertex $z \in V(P)$, the distance between two vertices in T' can be calculated as follows:

- If $x, y \in V(X_{>k}) \cup V(Y_{>k})$, we get $d_{T'}(x, y) = d_T(x, y)$.
- If $x, y \in V(T) \setminus (V(X_{>k}) \cup V(Y_{>k}))$, we have $d_{T'}(x, y) = d_T(x, y)$.
- If $x \in V(Z)$ and $y \in V(A_{>k})$ with $A = X, Y$, we obtain $d_{T'}(x, y) = d_T(x, y)$.
- If $x \in V(A_{>k})$ and $y \in V(A_i)$, we get $d_{T'}(x, y) = d_T(x, y) + 2i$ for $i = 1, \ldots, k$ and $A = X, Y$.
- If $x \in V(X_{>k})$ and $y \in V(Y_i)$ or $x \in V(Y_{>k})$ and $y \in V(X_i)$, $i = 1, \ldots, k$, we obtain $d_{T'}(x, y) = d_T(x, y) - 2i$.

Analogously we get

$$W(T') - W(T) = \sum_{i=1}^{k} 2i(|V(X_i)| - |V(Y_i)|)(|V(X_{>k})| - |V(Y_{>k})|) \leq 0,$$

which completes the proof. □

Lemma 3.17. *Let T be a tree with minimum Wiener index among all trees of the same degree sequence. Furthermore let P be a path of T with $|V(X_i)| \geq |V(Y_i)|$ for $i = 1, 2, \ldots, k-1$ and $|V(X_{>k})| \geq |V(Y_{>k})|$, $k = 1, \ldots, m-1$. Then we can assume*

$$|V(X_k)| \geq |V(Y_k)|.$$

Proof. Let us assume that $|V(X_k)| < |V(Y_k)|$. Furthermore let T' be the tree obtained from T by switching X_k and Y_k. We show that $W(T') \leq W(T)$ by the same case distinction as in the proof of Lemma 3.16. Thus we obtain for $z \notin V(P)$

$$\begin{aligned} W(T') - W(T) = &\sum_{i=1}^{k-1} (2i-1)(|V(X_i)| - |V(Y_i)|)(|V(X_k)| - |V(Y_k)|) \\ &+ (2k-1)(|V(X_{>k})| - |V(Y_{>k})|)(|V(X_k)| - |V(Y_k)|) \leq 0, \end{aligned}$$

and for $z \in V(P)$

$$\begin{aligned} W(T') - W(T) = &\sum_{i=1}^{k-1} 2i(|V(X_i)| - |V(Y_i)|)(|V(X_k)| - |V(Y_k)|) \\ &+ 2k(|V(X_{>k})| - |V(Y_{>k})|)(|V(X_k)| - |V(Y_k)|) \leq 0. \end{aligned}$$

□

Lemma 3.18. *Let T be a tree with minimum Wiener index among all trees of the same degree sequence. Furthermore let P be a path of T with $|V(X_i)| \geq |V(Y_i)|$ for $i = 1, 2, \ldots, k-1$ and $V(X_{>k-1})| \geq |V(Y_{>k-1})|$, $k = 1, \ldots, m$. Then we can assume*

$$\deg_T(x_k) \geq \deg_T(y_k).$$

Proof. For contradiction let us assume that $a = \deg_T(x_k) < \deg_T(y_k) = a+b$ with $a, b \geq 1$. Thus deleting y_k leads to $a+b$ connected components. We choose $C_1, \ldots, C_b$ to be such components that no vertex of $V(P)$ is one of their elements. Furthermore let $u_1, \ldots, u_b$ be the neighbours of y_k lying in $C_1, \ldots, C_b$ and let B be the set of all vertices in $C_1, \ldots, C_b$. Then we define T' to be the tree with vertex set $V(T)$ and edge set

$$(E(T) \setminus \{(y_k, u_i) \,:\, i = 1, \ldots, b\}) \cup \{(x_k, u_i) \,:\, i = 1, \ldots, b\}.$$

Thus T' and T have the same degree sequence and, similar to the previous proofs, we get for $z \notin V(P)$

$$\begin{aligned} W(T') - W(T) = &\sum_{i=1}^{k-1} (2i-1) \underbrace{|B|}_{>0} \underbrace{(|V(Y_i)| - |V(X_i)|)}_{\leq 0} \\ &+ (2k-1) \underbrace{|B|}_{>0} \underbrace{(|V(Y_{>k-1})| - |B| - |V(X_{>k-1})|)}_{\leq 0} \leq 0, \end{aligned}$$

and for $z \in V(P)$

$$\begin{aligned} W(T') - W(T) = &\sum_{i=1}^{k-1} 2i|B|(|V(Y_i)| - |V(X_i)|) \\ &+ 2k|B|(|V(Y_{>k-1})| - |B| - |V(X_{>k-1})|) \leq 0. \end{aligned}$$

Therefore in both cases the Wiener index of T' is at most as great as the Wiener index of T, although $\deg_{T'}(x_k) \geq \deg_{T'}(y_k)$. □

Remark 3.19. If at least one inequality in the conditions of Lemma 3.16, Lemma 3.17 or Lemma 3.18 is strict, the assumptions in the conclusions become forced.

Up to now we have considered arbitrary paths in a tree T which is an optimal solution to the Problem 3.5. Now, let P^* be a maximal path of T. If P^* has odd length $2m-1$, we label its vertices as $v_{l_1}, \dots, v_1, u_1, \dots, u_{l_2}$ in this order with $l_1 + l_2 = 2m$. If P^* has even length $2m$, we label them as $v_{l_1}, \dots, v_1, u_1, \dots, u_{l_2+1}$ in this order with $l_1 + l_2 + 1 = 2m + 1$.

Analogously we label the connected components of T which are obtained by deleting all edges of P as U_i and V_i, respectively. We choose the labelling such that U_1 is the component with most vertices.

Lemma 3.20. *Let T be a tree with minimum Wiener index among all trees of the same degree sequence. Furthermore let P^* be a maximal path of T. Then we can assume that*

$$|V(U_1)| \geq |V(V_1)| \geq |V(U_2)| \geq |V(V_2)| \geq \cdots \geq |V(U_m)| = |V(V_m)| = 1$$

for P^* having odd length $2m-1$ and

$$|V(U_1)| \geq |V(V_1)| \geq |V(U_2)| \geq |V(V_2)| \geq \cdots \geq |V(V_m)| = |V(U_{m+1})| = 1$$

for P^* having even length $2m$. In both cases we have $l_1 = l_2 = m$.

Proof. We show the lemma by induction. So, let P^* be of odd length. Since U_1 is the component with most vertices, we have $|V(U_1)| \geq |V(V_1)|$ and $|V(U_1)| \geq |V(U_2)|$. If $|V(U_2)| > |V(V_1)|$, we just need to switch U_2 and V_1 to get

$$|V(U_1)| \geq |V(V_1)| \geq |V(U_2)|.$$

Now suppose that we have

$$|V(U_1)| \geq |V(V_1)| \geq |V(U_2)| \geq |V(V_2)| \geq \cdots \geq |V(V_{k-1})| \geq |V(U_k)|$$

for some $k \leq \min\{l_1, l_2\}$. To show that we can assume $|V(U_k)| \geq |V(V_k)|$ we suppose that $|V(U_k)| < |V(V_k)|$ must hold. According to Lemma 3.16 applied to the labelling $x_i = u_i$ and $y_i = v_i$, $i = 1, \dots, k-1$, we have $|V(U_{>k-1})| \geq |V(V_{>k-1})|$. Together we obtain

$$|V(U_{>k})| = |V(U_{>k-1})| - |V(U_k)| > |V(V_{>k-1})| - |V(V_k)| = |V(V_{>k})|.$$

Thus we can apply Lemma 3.17 and we get $|V(U_k)| \geq |V(V_k)|$, which is a contradiction. Therefore we have

$$|V(U_1)| \geq |V(V_1)| \geq |V(U_2)| \geq |V(V_2)| \geq \cdots \geq |V(U_k)| \geq |V(V_k)|.$$

Analogously we show that $|V(V_k)| \geq |V(U_{k+1})|$ for $k < \min\{l_1, l_2\}$ by assuming that $|V(V_k)| < |V(U_{k+1})|$. Using the labelling $z = u_1$, $x_i = v_i$ and $y_i = u_{i+1}$, $i = 1, \dots, k-1$, we get

$$|V(V_{>k-1})| = |V(X_{>k-1})| \geq |V(Y_{>k-1})| = |V(U_{>k})|$$

by Lemma 3.16. If $|V(Y_k)| = |V(U_{k+1})| > |V(V_k)| = |V(X_k)|$ holds, we get

$$|V(U_{>k+1})| = |V(U_{>k})| - |V(U_{k+1})| > |V(V_{>k-1})| - |V(V_k)| = |V(V_{>k})|.$$

Applying Lemma 3.17 shows that $|V(V_k)| \geq |V(U_{k+1})|$ must hold, which is a contradiction. Thus we have

$$|V(U_1)| \geq |V(V_1)| \geq |V(U_2)| \geq |V(V_2)| \geq \cdots \geq |V(V_k)| \geq |V(U_{k+1})|.$$

Altogether we have

$$|V(U_1)| \geq |V(V_1) \geq \cdots \geq |V(U_k)| \geq |V(V_k)|$$

with $k = \min\{l_1, l_2\}$. If $k = l_2$, we obtain that $|V(U_{>k})| = 0$ since P^* is maximal. According to Lemma 3.16 we have $|V(V_{>k})| = 0$ and thus $l_1 = l_2 = m$. If $k = l_1 < l_2$ we analogously get that $|V(V_{>k})| = 0$. Using the labelling $z = u_1$, $x_i = v_i$ and $y_i = u_{i+1}$, $i = 1, \ldots, k$, we obtain $|V(U_{>k+1}| = 0$ by Lemma 3.16. Therefore we have $l_1 \leq l_2 \leq l_1 + 1$. In case $l_2 = l_1 + 1$, we get $l_1 + l_2 = 2l_1 + 1$, which is a contradiction to the assumption that P^* has odd length. Thus

$$l_1 = l_2 = m$$

and, since P^* is maximal,

$$|V(U_m)| = |V(V_m)| = 1.$$

If P^* has even length, the proof is analogous. □

Lemma 3.21. *Let T be a tree with minimum Wiener index among all trees of the same degree sequence. Furthermore let P^* be a maximal path of T with labelling such that Lemma 3.20 holds. Then we have*

$$\deg_T(u_1) \geq \deg_T(v_1) \geq \deg_T(u_2) \geq \deg_T(v_2) \geq \cdots \geq \deg_T(u_m) = \deg_T(v_m) = 1$$

for P^* having odd length $2m - 1$ and

$$\deg_T(u_1) \geq \deg_T(v_1) \geq \deg_T(u_2) \geq \deg_T(v_2) \geq \cdots \geq \deg_T(v_m) = \deg_T(u_{m+1}) = 1$$

for P^* having even length $2m$.

Proof. In case P^* has odd length, we have according to Lemma 3.20

$$|V(U_1)| \geq |V(V_1)| \geq |V(U_2)| \geq |V(V_2)| \geq \cdots \geq |V(U_m)| = |V(V_m)| = 1.$$

Thus we obtain

$$|V(U_{>i-1})| = \sum_{k=i}^{m} |V(U_k)| \geq \sum_{k=i}^{m} |V(V_k)| = |V(V_{>i-1})|$$

for $i = 1, \ldots, m$. Applying Lemma 3.18 to P^* in the setting that $x_i = u_i$ and $y_i = v_i$ for $i = 1, \ldots, m$ leads to $\deg_T(x_i) \geq \deg_T(y_i)$, which means

$$\deg_T(u_i) \geq \deg_T(v_i) \qquad i = 1, \ldots, m.$$

On the other hand we have

$$|V(V_{>i-1})| = \sum_{k=i}^{m} |V(V_k)| \geq \sum_{k=i+1}^{m} |V(U_k)| = |V(U_{>i})|$$

for $i = 1, \ldots, m$. Using the assignment $z = u_1$, $x_i = v_i$ and $y_i = u_{i+1}$ with $i = 1, \ldots, m-1$, we get $\deg_T(x_i) \geq \deg_T(y_i)$ by applying Lemma 3.18, which means

$$\deg_T(v_i) \geq \deg_T(u_{i+1}) \qquad i = 1, \ldots, m-1.$$

As $|V(U_m)| = |V(V_m)| = 1$, we further have

$$\deg_T(u_m) = \deg_T(v_m) = 1.$$

Altogether we obtain

$$\deg_T(u_1) \geq \deg_T(v_1) \geq \deg_T(u_2) \geq \deg_T(v_2) \geq \cdots \geq \deg_T(u_m) = \deg_T(v_m) = 1.$$

It is easy to see that the proof of the second case is analogous. □

Finally we once again need the concept of the centroid of a tree and one further characteristic of centroids which can be found in [24]:

Theorem 3.22. *Let T be a tree and $C(T)$ its centroid. Then we have $d_T(c) \leq d_T(v)$ for all $c \in C(T)$ and $v \in V(T)$.*

Proof. Let $v \in V(T)$ with $v \neq c$ and P be the path between v and c. Furthermore let $B_1, \ldots, B_k$ be the branches of c such that $v \in V(B_1)$. Let $l \geq 1$ be the length of P and let B'_i be the subtree of T induced by $V(B_i) \setminus \{c\}$. Then we have the following three cases:

- If $x \in V(\bigcup_{i=2}^{k} B'_i)$, we get $d_T(v, x) = d_T(c, x) + l$.
- For all $x \in V(P)$ the sums over their distances to v and c, respectively, cancel each other out.
- If $x \in V(B'_1) \setminus V(P)$, we have $d_T(c, x) \leq d_T(v, x) + l$.

If $|C(T)| = 1$, we have $\frac{n-1}{2} \geq bw(B_1) = |V(B'_1)|$ according to Theorem 3.10 and thus we get $\frac{n+1}{2} \leq |V(\bigcup_{i=2}^{k} B_i)| = |V(\bigcup_{i=2}^{k} B'_i)| + 1$. Therefore we have

$$|V(B'_1) \setminus V(P)| \leq \frac{n-1}{2} - l \leq \frac{n-3}{2} < \frac{n-1}{2} \leq |V(\bigcup_{i=2}^{k} B'_i)|$$

and we obtain

$$d_T(v) - d_T(c) \geq \sum_{x \in V(\bigcup\limits_{i=2}^{k} B'_i)} l - \sum_{x \in V(B'_1) \setminus V(P)} l > 0.$$

If $|C(T)| = 2$ and $v \in C(T)$, we have $\frac{n}{2} = bw(B_1) = |V(B'_1)|$ according to Theorem 3.10 and therefore we obtain $\frac{n}{2} = |V(\bigcup\limits_{i=2}^{k} B_i)| = |V(\bigcup\limits_{i=2}^{k} B'_i)| + 1$. Furthermore we have $l = 1$ and thus

$$d_T(v) = d_T(c).$$

If $|C(T)| = 2$, $v \notin C(T)$ and $C(T) \cap V(B_1) \neq \emptyset$, we again have $\frac{n}{2} = bw(B_1) = |V(B'_1)|$ according to Theorem 3.10 and therefore $\frac{n}{2} = |V(\bigcup\limits_{i=2}^{k} B_i)| = |V(\bigcup\limits_{i=2}^{k} B'_i)| + 1$. Furthermore in this case we have $l \geq 2$ and thus

$$d_T(v) - d_T(c) > 0.$$

If $|C(T)| = 2$, $v \notin C(T)$ and $C(T) \cap V(B_1) = \emptyset$, we get $\frac{n}{2} > bw(B_1) = |V(B'_1)|$ according to Theorem 3.10 and $\frac{n}{2} < |V(\bigcup\limits_{i=2}^{k} B_i)| = |V(\bigcup\limits_{i=2}^{k} B'_i)| + 1$, as n has to be even. With this and $l \geq 1$ we obtain

$$d_T(v) - d_T(c) > 0,$$

which completes the proof. □

Lemma 3.23. *Let T be a tree. Considering any maximal path P^* of T with labelling as in Lemma 3.20, then*

$$d_T(u_1) \leq d_T(x)$$

for all $x \in V(P^*)$.

Proof. If $u_1 \in C(T)$ everything is done according to Theorem 3.22. Thus let us assume that $u_1 \notin C(T)$.

Let $c \in C(T)$ with $c \notin V(P^*)$. Then it is obvious that there must be a branch B_1 of c such that $V(P^*) \subseteq V(B_1)$. As $|V(B_1)| \leq \frac{n}{2}$ according to Theorem 3.10 and U_1 is the component with most vertices, there must be a path between u_1 and c such that no other vertex of P^* lies on it. Thus $d_T(u_1) \leq d_T(x)$ for all $x \in V(P^*)$.

Now, let $C(T) \subseteq V(P^*)$ and $c \in C(T)$. Hence there exists a k such that $c = u_k$ or $c = v_k$. We consider the case that $c = u_k$, the other one is similar. As $u_1 \notin C(T)$, we have $c \neq u_1$. For the distance between u_1 and a vertex x we distinguish the following three cases:

- For $x \in V(V_{>0}) \cup V(U_1)$ we obtain $d_T(u_1, x) = d_T(u_k, x) - (k - 1)$.
- If $x \in V(U_{>k-1})$, we get $d_T(u_1, x) = d_T(u_k, x) + (k - 1)$.

- If $x \in V(U_i)$, $2 \le i \le k-1$, we have $d_T(u_1, x) = d_T(u_k, x) + (2i - k - 1)$.

Thus we obtain for $k \ge 2$

$$\begin{aligned}
d_T(u_1) - d_T(u_k) = & \sum_{x \in V(V_{>0}) \cup V(U_1)} -(k-1) + \sum_{x \in V(U_{>k-1})} (k-1) \\
& + \sum_{i=2}^{k-1} \sum_{x \in V(U_i)} (2i - k - 1) \\
= & -(k-1)(|V(V_{>0})| + |V(U_1)|) + (k-1)|V(U_{>k-1})| \\
& + \sum_{i=2}^{k-1} |V(U_i)|(2i - k - 1) \\
\le & -(k-1)|U_1| - (k-1) \sum_{i=1}^{k-2} |V(V_i)| \\
& + \sum_{i=2}^{k-1} |V(U_i)|(2i - k - 1) \\
\le & \, 0,
\end{aligned}$$

where the first inequality holds due to $|V(V_i)| \ge |V(U_{i+1})|$ and the second one holds as $k - 3 \ge 2i - k - 1$ for $i \le k-1$ and hence $-(k-1)|V(V_i)| + (2i - k + 1)|V(U_{i+1})| \le 0$. But this is a contradiction to the fact that $u_1 \notin C(T)$. Thus, according to Theorem 3.22, we obtain $d_T(u_1) \le d_T(x)$ for all $x \in V(P^*)$. $\square$

Now we can prove the following theorem given in [21] which settles Problem 3.5:

Theorem 3.24. *Let $(d_1, \dots, d_n)$ be an integer sequence with*

$$d_1 \ge d_2 \ge \cdots \ge d_k \ge 2 > 1 = d_{k+1} = \cdots = d_n$$

and

$$\sum_{i=1}^{n} d_i = 2(n-1).$$

Then the greedy tree with degree sequence $(d_1, \dots, d_n)$ minimizes the Wiener index among all trees with same degree sequence.

Proof. Let T be the tree which minimizes the Wiener index among all trees with degree sequence $(d_1, \dots, d_n)$. To show that T is the greedy tree, we show that T fulfils Lemma 3.15. We distinguish two cases: either $|C(T)| = 1$ or $|C(T)| = 2$. If there is only one vertex in the centroid, we choose it as root v. If there are two vertices in $C(T)$, we choose one as root v and the other one as its leftmost neighbour. According to Lemma 3.21 and Lemma 3.23 v is the vertex with largest degree in T. Thus 1. of Lemma 3.15 is satisfied.

Let l_1 and l_2 be two leaves of T. If v lies on the path between l_1 and l_2, we obtain $|d_T(l_1, v) - d_T(l_2, v)| \leq 1$ since $u_1 = v$ using the labelling of Lemma 3.20. Otherwise there is a vertex u lying on the path between l_1 and l_2 such that l_i is connected with v via u, $i = 1, 2$. As the path is of maximal length, we can apply Lemma 3.20 to it and obtain $u = u_1$ through similar considerations as made in Lemma 3.23. Thus we have

$$|d_T(l_1, v) - d_T(l_2, v)| = |d_T(l_1, u) - d_T(l_2, u)| \leq 1$$

and 2. of Lemma 3.15 is satisfied.

Let x and y be two vertices with $d_T(x, v) < d_T(y, v)$. If v, x and y are collinear, we just choose a path of maximal length such that $v, x, y \in V(P)$. It follows immediately from Lemma 3.21 that $\deg_T(x) \geq \deg_T(y)$. If v, x and y are not collinear, there exists a vertex u lying on the path P between x and y such that a is connected with v via u, $a = x, y$. We prolong P on both sides to gain a path P' of maximal length. Again we get that $u = u_1$ and, by using Lemma 3.20, we obtain that $x = u_{i+1}$, $y = v_j$ or $x = v_i$, $y = u_{j+1}$ with $i = d_T(x, v) - d_T(u, v)$ and $j = d_T(y, v) - d_T(u, v)$. Since $i < j$, we get $\deg_T(x) \geq \deg_T(y)$ according to Lemma 3.21. Hence 3. of Lemma 3.15 is satisfied.

Now, let x and y be two non-leaves with $d_T(x, v) = d_T(y, v)$ and $\deg_T(x) > \deg_T(y)$. Moreover let x' and y' be two further vertices with

$$d_T(x', v) = d_T(x', x) + d_T(x, v) = d_T(y', y) + d_T(y, v) = d_T(y', v).$$

Let us consider a path P with maximal length and $x, y, x', y' \in V(P)$. Applying Lemma 3.20 we obtain that there exists a vertex $u \in V(P)$ such that both x and y are connected with v via u and therefore $u = u_1$. Furthermore, as $\deg_T(x) > \deg_T(y)$, we get $x = v_i$, $y = u_{i+1}$, $x' = v_j$ and $y' = u_{j+1}$ with $i = d_T(x, u)$ and $j = d_T(x', u)$. Thus we have $\deg_T(x') \geq \deg_T(y')$ and 4. of Lemma 3.15 is satisfied.

Again, let x and y be two non-leaves with $d_T(x, v) = d_T(y, v)$ and $\deg_T(x) > \deg_T(y)$ and let u be the vertex lying on the path between x and y such that both x and y are connected with v via u. Furthermore let x' and y' be two vertices with $d_T(x', v) = d_T(y', v) = d_T(x, v)$ such that a' and a have a_0 as common neighbour, $a = x, y$. Moreover let x'' and y'' be two further vertices with

$$d_T(x'', v) = d_T(x'', x') + d_T(x', v) = d_T(y'', y') + d_T(y', v) = d_T(y'', v).$$

Since 4. of Lemma 3.15 holds, we have $\deg_T(x_0) \geq \deg_T(y_0)$. Therefore considering the path P of maximal length with $u, x_0, y_0, x', y', x'', y'' \in V(P)$, we again obtain $u = u_1$, $x_0 = v_i$, $y_0 = u_{i+1}$, $x' = v_{i+1}$, $y' = u_{i+2}$, $x'' = v_j$ and $y'' = u_{j+1}$ with $i = d_T(x_0, u)$ and $j = d_T(x'', u)$ by applying Lemma 3.20. According to Lemma 3.21 we get $\deg_T(x') \geq \deg_T(y')$ and $\deg_T(x'') \geq \deg_T(y'')$ and 5. of Lemma 3.15 is satisfied.

Finally, by Lemma 3.15, T is the greedy tree. □

Maximizing the Wiener index

Problem 3.7. Given an integer sequence $(d_1, \ldots, d_n)$ with

$$d_1 \geq d_2 \geq \cdots \geq d_k \geq 2 > 1 = d_{k+1} = \cdots = d_n$$

and

$$\sum_{i=1}^{n} d_i = 2(n-1),$$

we want to find a tree with degree sequence $(d_1, \ldots, d_n)$ which maximizes the Wiener index among all trees of the same degree sequence.

Since it is an open problem whether or not finding an optimal solution is NP-hard, we can only give some characteristics of the optimal solution. The first of these characteristics, which reduces the set of possible solutions to the set of caterpillars, was published by Shi in 1993 (see [19]).

Lemma 3.25. *Let $(d_1, d_2, \ldots, d_n)$ be a degree sequence with $\sum_{i=1}^{n} d_i = 2(n-1)$. Furthermore let $T_{\max}$ be the tree with maximal Wiener index over all trees which have this particular degree sequence. Then $T_{\max}$ is a caterpillar.*

Proof. We prove this theorem by contradiction. So, let us assume that $T_{\max}$ is not a caterpillar. Then there exists a path of length $l \geq 4$.

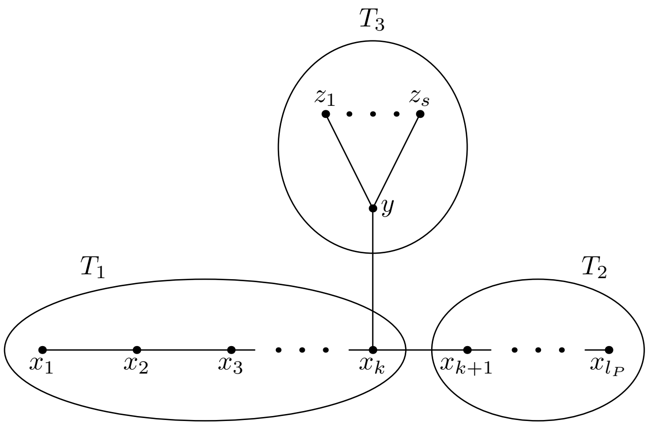


Figure 3.4

Let $P = (x_1, x_2, \ldots, x_{l_P})$ be a longest path of $T_{\max}$. As $T_{\max}$ is not a caterpillar, there exists an x_k, $2 < k < l_P - 1$, such that x_k is the endpoint of a path P_1 not containing any other vertex of P and having length $l_{P_1} \geq 2$. We denote the vertex connected with x_k on P_1 by y. Then let $N(y) = \{x_k, z_1, \ldots, z_s\}$, $s \geq 1$, be the neighbours of y. Furthermore let T_1 be the subtree of $T_{\max}$ after deleting the edges (x_k, x_{k+1}) and (x_k, y) containing x_k. By the same operations we get T_2 as the subtree containing x_{k+1} and T_3 as the one containing y, as illustrated in Figure 3.4. Without loss of generality we can also assume that $|V(T_1)| > |V(T_2)|$.

Now we are going to construct a tree $T' \neq T_{\max}$, such that $W(T') > W(T)$. In order to do so we replace each edge (y, z_i) of $T_{\max}$ by the new edge (x_{l_P}, z_i), $i = 1, \ldots, s$. Hence we obtain

$$\deg_{T'}(y) = 1 = \deg_{T_{\max}}(x_{l_P})$$

$$\begin{aligned}\deg_{T'}(x_{l_P}) &= \deg_{T_{\max}}(y)\\ \deg_{T'}(z_i) &= \deg_{T_{\max}}(z_i)\end{aligned}$$

with $i = 1, \dots, s$, and therefore the degree sequence of T' and $T_{\max}$ is the same.

Now we consider the distance between two vertices. It is easy to see that the distance within $T_{\max}$ and T' is the same in the following five cases:

- $u,\, v \in V(T_1)$
- $u,\, v \in V(T_2)$
- $u,\, v \in V(T_3) \setminus \{y\}$
- $u \in V(T_1),\, v \in V(T_2)$
- $u \in V(T_1) \cup V(T_2)$ and y.

Furthermore, as for the degree, x_{l_P} takes the place of y and vice versa, concerning the distance to a vertex in $V(T_3) \setminus \{y\}$. The only two cases in which the distance really differs are the ones where $u \in V(T_3) \setminus \{y\}$ and $v \in V(T_1)$ or $v \in V(T_2) \setminus \{x_{l_P}\}$. In case $v \in V(T_1)$ we obtain

$$\begin{aligned}d_{T_{\max}}(u, v) &= d_{T_{\max}}(u, y) + 1 + d_{T_{\max}}(x_k, v)\\ d_{T'}(u, v) &= d_{T_{\max}}(u, y) + d_{T_{\max}}(x_{l_P}, x_k) + d_{T_{\max}}(x_k, v),\end{aligned}$$

and in case $v \in V(T_2) \setminus \{x_{l_P}\}$ we get

$$\begin{aligned}d_{T_{\max}}(u, v) &= d_{T_{\max}}(u, y) + 1 + d_{T_{\max}}(x_k, v)\\ d_{T'}(u, v) &= d_{T_{\max}}(u, y) + d_{T_{\max}}(x_{l_P}, v).\end{aligned}$$

Since, at the beginning of the proof, we assumed that P is a longest path in $T_{\max}$, we obtain that $d_{T_{\max}}(x_k, v) \leq d_{T_{\max}}(x_k, x_{l_P})$ for all $v \in V(T_2)$.

Thus we get

$$\begin{aligned}W(T') - W(T_{\max}) = &\sum_{u \in V(T_3)\setminus\{y\}} \Bigg[\sum_{v \in V(T_1)} (d_{T_{\max}}(x_k, x_{l_P}) - 1)\\ &+ \sum_{v \in V(T_2)\setminus\{x_{l_P}\}} (d_{T_{\max}}(v, x_{l_P}) - d_{T_{\max}}(x_k, v) - 1)\Bigg]\\ > &\sum_{u \in V(T_3)\setminus\{y\}} \Bigg[\sum_{v \in V(T_2)\setminus\{x_{l_P}\}} d_{T_{\max}}(v, x_{l_P})\\ &+ \underbrace{d_{\max}(x_k, x_{l_P})}_{\geq 2}(|V(T_1)| - |V(T_2)| + 1)\\ &- (|V(T_1)| + |V(T_2)| - 1)\Bigg]\end{aligned}$$

$$\begin{aligned}
&\geq \sum_{u\in V(T_3)\setminus\{y\}} \Bigg[\sum_{v\in V(T_2)\setminus\{x_{l_P}\}} d_{T_{\max}}(v, x_{l_P}) \\
&\qquad + \underbrace{|V(T_1)| - (|V(T_2)| - 1)}_{\geq 2 \text{ due to assumption}} - 2(|V(T_2)| - 1)\Bigg] \\
&\geq 0,
\end{aligned}$$

where the last inequality holds as $d_{T_{\max}}(v, x_{l_P}) \geq 2$ for all $v \in V(T_2) \setminus \{x_{l_P-1}, x_{l_P}\}$. But this is a contradiction to the condition that $T_{\max}$ has maximal Wiener index. □

Now, by using Lemma 3.25 we are also able to solve Problem 3.4:

Theorem 3.26. *The tree $T_{\max}$ maximizes the Wiener index among all trees of order n, $n \equiv 2 \mod (\Delta - 1)$, with all vertices of either degree 1 or Δ if and only if $T_{\max}$ is a caterpillar.*

Proof. Let $(\underbrace{\Delta, \dots, \Delta}_{k \text{ times}}, \underbrace{1, \dots, 1}_{n-k \text{ times}})$ be the degree sequence of T with some k. Then

$$\sum_{i=1}^{n} \deg_T(v_i) = k\Delta + n - k = 2(n-1).$$

But this is just equivalent to

$$n \equiv 2 \mod (\Delta - 1),$$

which means that a tree T with order n and all vertices being either leaves or of degree Δ exists and, since the number of possible trees is finite, also a tree $T_{\max}$ with maximum Wiener index exists. According to Lemma 3.25 $T_{\max}$ is a caterpillar.

On the other hand let T be a caterpillar. If $T_{\max}$ is an optimal solution to Problem 3.4, we have already shown that $T_{\max}$ is a caterpillar. As the caterpillar with vertices of either degree 1 or Δ is unique, we get $T = T_{\max}$, which completes the proof. □

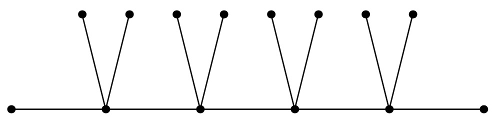


Figure 3.5: A caterpillar with $n = 14$ and $\Delta = 4$.

Example 3.27. Let $n = 14$ and $\Delta = 4$. Then $n \equiv 2 \mod (\Delta - 1)$ and the tree with maximum Wiener index among all trees with vertex degree either 1 or Δ is the caterpillar shown in Figure 3.5.

As we have seen in Lemma 3.25, the optimal solution of Problem 3.7 is a caterpillar in all cases. Therefore, in the following we show a formula for computing the Wiener index of caterpillars and discuss how to arrange the vertices on the longest path of the caterpillar in order to maximize its Wiener index (see [25]).

Lemma 3.28. *Let T be a caterpillar on n vertices and $(\deg_T(v_1), \deg_T(v_2), \dots, \deg_T(v_n))$ its unordered degree sequence, where $\deg_T(v_i) \geq 2$, $1 \leq i \leq k$, belongs to the i-th vertex of the path formed by all k non-leaves. Then*

$$W(T) = (n-1)^2 + \sum_{i=1}^{k-1} \Bigg(\sum_{j=1}^{i} (\deg_T(v_j) - 1) \Bigg) \Bigg(\sum_{j=i+1}^{k} (\deg_T(v_j) - 1) \Bigg).$$

Proof. In order to use formula (2.1) we have to calculate the number of vertices of the two connected components of T after deleting $e_i = (v_i, v_{i+1})$, $i = 1, \dots, k-1$:

$$\begin{aligned} n_{v_i}(e_i) &= \sum_{j=1}^{i} (\deg_T(v_j) - 1) + 1 \\ n_{v_{i+1}}(e_i) &= \sum_{j=i+1}^{k} (\deg_T(v_j) - 1) + 1. \end{aligned}$$

Furthermore we obtain for each of the remaining $(n-k)$ edges which have a leaf as pendant vertex that one of its subtrees contains one vertex and the other one $n-1$ vertices.

Thus we can easily calculate

$$\begin{aligned} W(T) = &\sum_{e=(u,v) \in E(T)} n_u(e) n_v(e) \\ &= (n-1)(n-k) + \sum_{i=1}^{k-1} \Bigg(\sum_{j=1}^{i} (\deg_T(v_j) - 1) + 1 \Bigg) \Bigg(\sum_{j=i+1}^{k} (\deg_T(v_j) - 1) + 1 \Bigg) \\ &= (n-1)(n-k) + (k-1) \Bigg(1 + \sum_{j=1}^{k} (\deg_T(v_j) - 1) \Bigg) \\ &\quad + \sum_{i=1}^{k-1} \Bigg(\sum_{j=1}^{i} (\deg_T(v_j) - 1) \Bigg) \Bigg(\sum_{j=i+1}^{k} (\deg_T(v_j) - 1) \Bigg) \\ &= (n-1)^2 + \sum_{i=1}^{k-1} \Bigg(\sum_{j=1}^{i} (\deg_T(v_j) - 1) \Bigg) \Bigg(\sum_{j=i+1}^{k} (\deg_T(v_j) - 1) \Bigg), \end{aligned}$$

where the last equation holds since $\sum_{j=1}^{k} (\deg_T(v_j) - 1) + 1 = n - 1$. $\square$

Now, we can characterize the tree with maximum Wiener index by the following theorem:

Theorem 3.29. *Let T be a tree and $(d_1, d_2, \dots, d_n)$ its degree sequence with $d_i \geq 2$ for $i = 1, \dots, k$ and $d_{k+1} = \dots = d_n = 1$. Then T has the maximum Wiener index of all trees with the same degree sequence if and only if T is a caterpillar with $(d'_1, \dots, d'_k)$ being the permutation of $(d_1, \dots, d_k)$ such that d'_j is the degree of the j-th vertex on the path of non-leaves in T and $(d'_1, \dots, d'_k)$ maximizes*

$$\sum_{i=1}^{k-1}\Big(\sum_{j=1}^{i}(d''_j - 1)\Big)\Big(\sum_{j=i+1}^{k}(d''_j - 1)\Big),$$

where $(d''_1, \dots, d''_k)$ is a permutation of $(d_1, \dots, d_k)$.

Proof. If T is the tree with maximum Wiener index of all trees with the same degree sequence, we already know from Lemma 3.25 that T must be a caterpillar, and, from Lemma 3.28, that it must maximize the sum above.

Now, let T be a caterpillar with $\deg_T(v_j) = d'_j$, $j = 1, \dots, k$, in which v_j is the j-th vertex on the path of non-leaves and $(d'_1, \dots, d'_k)$ maximizes the sum above. Furthermore let T_1 be an arbitrary tree which fulfils the given degree sequence. Due to Lemma 3.25 there exists a caterpillar T_2 with the same degree sequence, $(\tilde{d}_1, \dots, \tilde{d}_k)$ being its permutation of $(d_1, \dots, d_k)$ and $W(T_1) \leq W(T_2)$. Because of Lemma 3.28 and the choice of T we also obtain the inequality

$$\begin{aligned} W(T_2) &= (n-1)^2 + \sum_{i=1}^{k-1}\Big(\sum_{j=1}^{i}(\tilde{d}_j - 1)\Big)\Big(\sum_{j=i+1}^{k}(\tilde{d}_j - 1)\Big) \\ &\leq (n-1)^2 + \sum_{i=1}^{k-1}\Big(\sum_{j=1}^{i}(d'_j - 1)\Big)\Big(\sum_{j=i+1}^{k}(d'_j - 1)\Big) = W(T), \end{aligned}$$

and therefore $W(T_1) \leq W(T)$, which completes the proof. $\square$

Another characteristic of the optimal tree, which is a very important criterion for eliminating candidates for optimal trees, is given in the following lemma:

Lemma 3.30. *Let $x_1 \geq x_2 \geq \dots \geq x_k \geq 1$ be integers with $k \geq 5$. Further let S_k be the set of all permutations of $\{1, \dots, k\}$ and suppose that $(y_1, \dots, y_k)$ is a permutation in S_k such that w.l.o.g. $y_1 \geq y_k$ and*

$$\sum_{i=1}^{k-1}\Big(\sum_{j=1}^{i} y_j\Big)\Big(\sum_{j=i+1}^{k} y_j\Big) = \max_{\pi \in S_k} \sum_{i=1}^{k-1}\Big(\sum_{j=1}^{i} x_{\pi(j)}\Big)\Big(\sum_{j=i+1}^{k} x_{\pi(j)}\Big).$$

Then there exists a $2 \leq t \leq k-2$ such that

$$y_1 \geq y_2 \geq \dots \geq y_{t-1} \geq y_t \leq y_{t+1} \leq \dots \leq y_k.$$

Proof. We define the function

$$f(s) = \sum_{i=1}^{s-2} y_i - \sum_{i=s+1}^{k} y_i$$

for $2 \leq s \leq k-1$. Obviously

$$f(2) = -\sum_{i=3}^{k} y_i < 0$$

and

$$f(k-1) = \sum_{i=1}^{k-3} y_i - y_k > 0,$$

since $y_1 \geq y_k$. Besides, f is strictly increasing. Therefore there exists a $2 \leq t_1 \leq k-2$ such that

$$f(t_1) \leq 0 \quad \text{and} \quad f(t_1+1) > 0.$$

This means that

$$\sum_{j=1}^{t_1-2} y_j \leq \sum_{j=t_1+1}^{k} y_j \tag{3.1}$$

$$\sum_{j=1}^{t_1-1} y_j > \sum_{j=t_1+2}^{k} y_j. \tag{3.2}$$

Now, let us consider the permutation $(z_1, \ldots, z_k) = (y_1, y_2, \ldots, y_{i-1}, y_{i+1}, y_i, y_{i+2}, \ldots, y_k)$. Since $(y_1, \ldots, y_k)$ is the optimal permutation, we obtain

$$\begin{aligned}
&\sum_{i=1}^{k-1}\Big(\sum_{j=1}^{i} y_j\Big)\Big(\sum_{j=i+1}^{k} y_j\Big) - \sum_{i=1}^{k-1}\Big(\sum_{j=1}^{i} z_j\Big)\Big(\sum_{j=i+1}^{k} z_j\Big) \\
&\quad = \Big(\sum_{j=1}^{i-1} y_j + y_i\Big)\Big(\sum_{j=i+2}^{k} y_j + y_{i+1}\Big) - \Big(\sum_{j=1}^{i-1} y_j + y_{i+1}\Big)\Big(\sum_{j=i+2}^{k} y_j + y_i\Big) \\
&\quad = (y_{i+1} - y_i)\Big(\sum_{j=1}^{i-1} y_j - \sum_{j=i+2}^{k} y_j\Big) \geq 0.
\end{aligned}$$

Due to equation (3.1) and the fact that f is strictly increasing we get

$$\sum_{j=1}^{i-1} y_j < \sum_{j=i+2}^{k} y_j \qquad \text{for } 1 \leq i \leq t_1 - 2,$$

and because of equation (3.2) we have

$$\sum_{j=1}^{i-1} y_j > \sum_{j=i+2}^{k} y_j \qquad \text{for } t_1 \leq i \leq k-1.$$

Thus altogether we obtain

$$\begin{aligned} y_{i+1} - y_i &\le 0 \qquad \text{for } 1 \le i \le t_1 - 2 \\ y_{i+1} - y_i &\ge 0 \qquad \text{for } t_1 \le i \le k - 1, \end{aligned}$$

which means

$$y_1 \ge \cdots \ge y_{t_1 - 1} \qquad \text{and} \qquad y_{t_1} \le \cdots \le y_k.$$

If $t_1 = 2$, then $f(2) < 0$ and therefore $y_1 \ge y_2$ and $t = 2$.

If $t_1 \ge 3$, we obtain either $y_{t_1-1} \ge y_{t_1}$ and $t = t_1$ or $y_{t_1-1} \le y_{t_1}$ and $t = t_1 - 1$. □

Remark 3.31. Note that for $k \le 3$ the tree T with maximum Wiener index is just the greedy caterpillar, which means that in this case there is also a t like in Lemma 3.30. A caterpillar is called *greedy* if the i-th vertex on the path of non-leaves has degree d_{2i-1} for $i \le \frac{k+1}{2}$ and degree $d_{2(k+1-i)}$ for $i > \frac{k+1}{2}$. For $k = 4$, the degree permutation (d_1, d_4, d_3, d_2) gives the optimal solution.

Proof. Case $k = 1, 2$: It is obvious.

Case $k = 3$: Let T_1 be the optimal caterpillar with degree sequence (y_1, y_2, y_3), such that w.l.o.g. $y_1 \ge y_3$ and $\deg_{T_1}(v_i) = y_i$ with v_i being the i-th vertex of the path formed by the non-leaves of T_1. Furthermore let T_2 be the caterpillar with degree sequence (y_1, y_3, y_2). Then

$$0 \le W(T_1) - W(T_2) = (y_1 + y_2)y_3 - (y_1 + y_3)y_2 = y_1(y_3 - y_2).$$

Thus we obtain $y_3 \ge y_2$ and altogether we get $y_1 \ge y_3 \ge y_2$, which leads to

$$(y_1, y_2, y_3) = (d_1, d_3, d_2).$$

Case $k = 4$: Let T_1 again be the optimal caterpillar with degree sequence (y_1, y_2, y_3, y_4) analogously. Furthermore let T_2 be the caterpillar with degree sequence (y_1, y_3, y_2, y_4) and T_3 with (y_1, y_2, y_4, y_3). Then

$$0 \le W(T_1) - W(T_2) = (y_1 + y_2)(y_3 + y_4) - (y_1 + y_3)(y_2 + y_4) = (y_1 - y_4)(y_3 - y_2)$$

and

$$0 \le W(T_1) - W(T_3) = (y_1 + y_2 + y_3)y_4 - (y_1 + y_2 + y_4)y_3 = (y_1 + y_2)(y_4 - y_3).$$

Therefore we obtain $y_3 \ge y_2$ and $y_4 \ge y_3$, and together with the assumption $y_1 \ge y_4$ we get

$$(y_1, y_2, y_3, y_4) = (d_1, d_4, d_3, d_2).$$

□

This reduces the number of different possible trees to quite a high extent, which leads to the algorithm described below.

Algorithm for finding the optimal solution

The input is an unsorted list of all degrees $d_i \geq 2$, $i = 1, \ldots, k$. Then the algorithm works as follows:

1. First the input needs to be prepared: sort the given degree sequence decreasingly, collect all equal degrees and save them as a list of lists l, in which the first number of each sub-list is the degree -1 and the second number tells how often the degree occurs.

2. In case the degree sequence is constant, there is nothing left to do but to compute the Wiener index and return it.

3. Otherwise the optimal solution is constructed in a recursive manner by filling up the path from both ends. To do so, construct a left list l_l and a right one l_r. Due to Lemma 3.30 the largest degree has to belong to the leftmost or rightmost vertex. W.l.o.g. assign it to the leftmost vertex, delete it from l and invoke the recursive function:

   (a) Take the first element (d_i, n_i) from l, where d_i is the biggest degree not assigned yet, and delete it.

   (b) For all possible $a, b \in \mathbb{N}_0$ such that $a + b = n_i$ assign a times d_i to l_l and b times d_i to l_r, such that the new elements are the rightmost of l_l and the leftmost of l_r. Go back to (3a) until l contains only one element.

   (c) Join l_l, the not yet assigned degrees and l_r in this order.

   (d) Compare the Wiener index of the currently constructed caterpillar and the largest Wiener index found so far. If the Wiener index of the current caterpillar is greater, save it as the new largest Wiener index found so far.

Since the recursion is handled in a depth-first manner, all possible caterpillars are constructed and compared.

Let $d_1 = \cdots = d_{k_1} > d_{k_1+1} = \cdots = d_{k_1+k_2} > \cdots > d_{k_1+\cdots+k_l+1} = \cdots = d_k \geq 2$. Then the number of the different caterpillars to consider is $k_1(k_2+1)\cdots(k_l+1)$, since the leftmost vertex has always degree d_1 and the last block of equal degrees is just assigned to the last free positions. The worst case occurs if all degrees are pairwise different, which means that there are 2^{n-2} different caterpillars to consider.

Although this algorithm finds the optimal solution, it is not very satisfactory because of its running time. Therefore a polynomial heuristic which may find a caterpillar with Wiener index equal or close to the Wiener index of the optimal solution is of interest.

A heuristic

The input once more is an unsorted list of all degrees $d_i \geq 2$, $i = 1, \ldots, k$. The main idea is to arrange the vertices such that they are concentrated equally at both ends of the path. Then the algorithm works as follows:

1. Sort the given degree sequence -1 decreasingly and save it as list l.

2. In case the degree sequence is constant, there is nothing left to do but to compute the Wiener index and return it.

3. Otherwise construct a left list l_l and a right one l_r. Take the first (second) element of l and assign it to the leftmost (rightmost) vertex which means to save it as leftmost (rightmost) element in l_l (l_r).

   Take the biggest non-assigned element of l and save it as the rightmost element of l_l if the sum over all elements of l_l is smaller than the sum over all elements of l_r. Otherwise save it as the leftmost element of l_r. Continue until all elements of l are assigned.

4. Join l_l and l_r in this order and save it as l_1.

5. Compute the number of vertices of the left (right) half of the so far constructed caterpillar and save it as sum_l (sum_r).

   In case $sum_l = sum_r$, there is nothing left to do but to compute the Wiener index and return it together with the caterpillar.

   Otherwise start with the rightmost vertex of the left half and the leftmost one of the right half and compare the vertex pairs until one end is reached: swap a vertex u belonging to the heavier side and a vertex v belonging to the other side if

   - Lemma 3.30 is not violated,
   - the position of v is at least as near to its closer end as the position of u to its closer end,
   - the difference $|sum_l - sum_r|$ does not increase.

   Compute this changing of positions once allowing that $|sum_l - sum_r|$ can stay the same after each changing of position and saving it as l_2, and once demanding the decrease of $|sum_l - sum_r|$ after each changing of position and save it as l_3.

6. Compare the Wiener index of l_1, l_2 and l_3 and return their maximum.

Because of the running time of the first algorithm I was only able to test the heuristic for smaller values of k, in particular I did so for $k \leq 30$. In order to create degree sequences for testing I used the random number generator of Mathematica 7.0. Furthermore I only considered uniform distributed numbers.

The heuristic seems to find either an optimal solution or a solution whose Wiener index is quite close to the Wiener index of the optimal solution. In particular the results of the testing are:

- The smaller the length of the interval from which the degrees were chosen the likelier it was that the heuristic found an optimal solution.
- When the interval length was about 100, the heuristic did not find the optimal solution in almost half the cases.
- When the heuristic did not find an optimal solution, the mean percentage error was smaller than 0.007%.
- The mean percentage error decreased for greater values of k or when the length of the interval from which the degrees were chosen increased.

An example for not finding the optimal solution is the sequence

$$(15, 14, 13, 12, 12, 9, 9, 8, 8, 7, 4, 3).$$

The sequence returned by the heuristic is $H = (15, 12, 9, 9, 8, 4, 3, 7, 8, 12, 13, 14)$ which has Wiener index 34329, whereas an optimal solution is $OS = (15, 13, 9, 9, 7, 4, 3, 8, 8, 12, 12, 14)$ with Wiener index 34333. Notice that the sum over all vertices in the left half of H (resp. OS) equals the sum over all vertices in the right half of H (resp. OS).

Chapter 4

Inverse problems - forbidden values

In this chapter we are going to deal with a more elementary question: which natural numbers are Wiener indices? In more detail the problem can be stated as follows:

Problem 4.1. Given a $w \in \mathbb{N}$, we want to find a graph G from a certain class, such that its Wiener index fulfils

$$W(G) = w.$$

As mentioned before, some chemical characteristics of a substance and the Wiener index of its molecular graph correlate. Thus it is of some interest to find a graph from a certain class with a particular Wiener index.

Over the years Problem 4.1 has been considered for different types of graphs. In the following sections we are going to study the solutions for some special graph classes.

4.1 Connected graphs

In case G is only a connected graph, a solution of Problem 4.1 was given by Gutman et al. in 1994 (see [12]).

Lemma 4.1. *Let G be a connected graph with $|V(G)| = n$ and $|E(G)| = m$. Furthermore let the length of all paths in G be less than three. Then*

$$W(G) = n(n-1) - m.$$

Proof. Since G consists of m edges, there are m pairs of vertices having distance one. Furthermore we assumed that the length of all paths is less than three. Thus the remaining $\binom{n}{2} - m$ vertex pairs are at distance two. Together we obtain

$$W(G) = 1 \cdot m + 2 \cdot \left[\binom{n}{2} - m\right] = n(n-1) - m.$$

$\square$

As a result we can state the following lemma.

Lemma 4.2. *For every integer value w, such that $\binom{n}{2} \leq w \leq (n-1)^2$, $n \geq 1$, there exists a connected graph G of order n and diameter $d(G) \leq 2$, such that $W(G) = w$.*

Proof. Let $m = |E(G)|$. As G is connected, we have $m \geq n-1$. It is obvious that the only graph with $m = n-1$ and diameter less than three is the star S_n, which has Wiener index $(n-1)^2$. If we add k further edges to S_n, we obtain $m = n-1+k$ and the length of each path still is less than three. According to Lemma 4.1, the Wiener index of this graph is $(n-1)^2 - k$. The maximal value of k is $\binom{n}{2} - (n-1)$, which is reached for the complete graph K_n. Thus for every m, $n-1 \leq m \leq \binom{n}{2}$, there exists a connected graph of order n with maximum path-length less than three and Wiener index

$$n(n-1) - \binom{n}{2} = \binom{n}{2} \leq W(G) \leq (n-1)^2 = n(n-1) - (n-1).$$

$\square$

Theorem 4.3. *Let $w \in \mathbb{N}_0$. Then there exists a connected graph G with $W(G) = w$ if and only if $w \in \mathbb{N}_0 \setminus \{2, 5\}$.*

Proof. The main idea of the proof is to show that for $n \geq 4$ the upper bound of the Wiener index of a graph on n vertices is smaller than the lower bound of the Wiener index of a graph on $n+1$ vertices in Lemma 4.2 by at most 1. This means that the inequality $(n-1)^2 \geq \binom{n+1}{2} - 1$ holds for $n \geq 4$. This can be easily seen since

$$(n-1)^2 - \binom{n+1}{2} + 1 = \frac{1}{2}(n^2 - 5n + 4) = \frac{1}{2}(n-4)(n-1) \geq 0$$

for $n \geq 4$. Thus we have that all $w \geq \binom{4}{2} = 6$ are Wiener indices of some connected graphs. Applying Lemma 4.2 to $n = 1, 2, 3$ we obtain that also 0, 1, 3 and 4 are Wiener indices.

Therefore it only remains to be shown that there is also no graph with diameter greater than two, which has Wiener index 2 or 5. Let G be a connected graph and P a path of G with length greater than two. This means that $|V(P)| \geq 4$ and hence $W(P) \geq 10$. Since P is a sub-graph of G, we obtain $W(G) \geq 10$, which completes the proof. $\square$

4.2 Bipartite graphs

Definition 4.2. Let G be a graph with vertex set $V(G)$ and edge set $E(G)$. Then G is called *bipartite* if there exist two subsets $V_1(G)$ and $V_2(G)$ such that $V(G) = V_1(G) \uplus V_2(G)$ and all edges in $E(G)$ have exactly one endvertex in $V_1(G)$ and one in $V_2(G)$.

Furthermore let $|V_1| = a$ and $|V_2| = b$. Then G is called a *bipartite graph on $a+b$ vertices.*

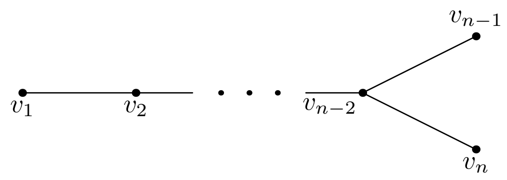


Figure 4.1: Q_n.

In 1995, Gutman and Yeh published a solution to Problem 4.1 for bipartite graphs (see [11]). They showed that all but 14 numbers are Wiener indices of bipartite graphs. In order to understand their proof, we have to consider some special graphs.

Let us denote the tree of order n which arises from the path P_{n-2} by connecting two further vertices as leaves to one of the two pendant vertices of P_{n-2} by Q_n (see Figure 4.1).

Furthermore we denote the circuit with n vertices by C_n and the complete bipartite graph on $a+b$ vertices by $K_{a,b}$.

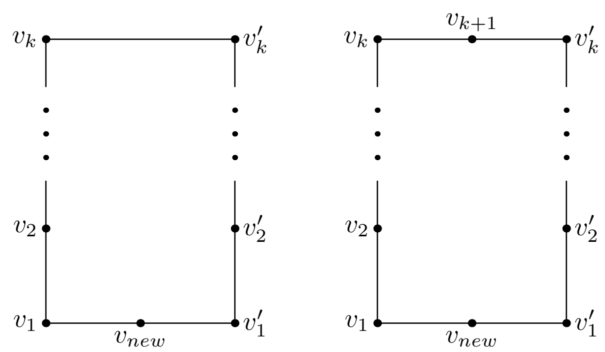


Figure 4.2: C_{n+1} arising from C_n with n even on the left and with n odd on the right.

Lemma 4.4. *The Wiener index of a circuit C_n of order n can be computed as*

$$W(C_n) = \begin{cases} \frac{n^3}{8} & \textit{if } n \textit{ is even,} \\ \frac{n^3-n}{8} & \textit{if } n \textit{ is odd.} \end{cases}$$

Proof. We show this formula with induction. It is easy to see that $W(C_3) = 3 = \frac{3^3-3}{4}$ and $W(C_4) = 8 = \frac{4^3}{8}$. So let the formula be true for $n \leq 2k$, $k \in \mathbb{N} \setminus \{1\}$. Then C_{2k+1} arises from C_{2k} by adding vertex v_{new} to C_{2k} as shown in Figure 4.2. Thus, using the labelling from Figure 4.2 we distinguish the following cases:

- For v_i and v_j with $1 \leq i < j \leq k$ the distance in C_{2k+1} and C_{2k} is the same. Analogously for v'_i and v'_j.
- For v_{new} and v_i, $i = 1, \ldots, k$, we have $d_{C_{2k+1}}(v_{new}, v_i) = i$. Analogously for v_{new} and v'_i.

- For v_i and v'_j with $1 \leq i \leq k$ and $k \geq j \geq k+1-i$ the distance in C_{2k+1} and C_{2k} is the same.
- For v_i and v'_j with $1 \leq i \leq k$ and $1 \leq j \leq k-i$ we have $d_{C_{2k+1}}(v_i, v'_j) = d_{C_{2k}}(v_i, v'_j)+1$.

Therefore we obtain for $n = 2k$

$$\begin{aligned}
W(C_{n+1}) &= \frac{n^3}{8} + 2\sum_{i=1}^{k} i + \sum_{i=1}^{k-1} i \\
&= \frac{n^3}{8} + k(k+1) + \frac{k(k-1)}{2} \\
&= \frac{n^3}{8} + \frac{3k^2+k}{2} = \frac{n^3}{8} + \frac{3n^2+2n}{8} \\
&= \frac{(n+1)^3-(n+1)}{8}.
\end{aligned}$$

With similar considerations we get for $n = 2k+1$ (see Figure 4.2)

$$\begin{aligned}
W(C_{n+1}) &= \frac{n^3-n}{8} + 2\sum_{i=1}^{k} i + k + 1 + \sum_{i=1}^{k} i \\
&= \frac{n^3-n}{8} + k(k+1) + k + 1 + \frac{k(k+1)}{2} \\
&= \frac{n^3-n}{8} + \frac{3k^2+5k+2}{2} = \frac{n^3-n}{8} + \frac{3n^2+4n+1}{8} \\
&= \frac{(n+1)^3}{8}.
\end{aligned}$$

□

Lemma 4.5. *The tree Q_n maximizes the Wiener index among all trees of order $n \geq 4$ that are different from P_n.*

Proof. According to Theorem 2.9 we have to maximize

$$\binom{n+1}{3} - \sum_{v \in V(T)} \sum_{1 \leq i < j < k \leq \deg_T(v)} |V_i|\,|V_j|\,|V_k| \tag{4.1}$$

over all trees T different from P_n. Obviously the first sum goes over all branching points of T and, since T is different from P_n, there must be at least one branching point. Thus, in order to maximize (4.1), T has to contain exactly one branching point. Let v be the branching point of T and V_l, $1 \leq l \leq \deg_T(v)$, the vertex sets of the subtrees gained by deleting v. Then we have to fulfil the condition

$$|V_1| + |V_2| + \cdots + |V_{\deg_T(v)}| = n - 1$$

for $|V_l| \geq 1$ with $1 \leq l \leq \deg_T(v)$. Thus there exists a V_i such that $|V_i| \geq \frac{n-1}{\deg_T(v)}$. Evidently $\deg_T(v) \leq n-1$. In case $\deg_T(v) = 3$, it is easy to see that the maximum in (4.1) is obtained by $|V_1| = n-3$, $|V_2| = |V_3| = 1$, which means that $T = Q_n$ and

$$|V_1|\,|V_2|\,|V_3| = n-3.$$

If $\deg_T(v) \geq 5$, we obtain

$$\begin{aligned}\sum_{1\leq i<j<k\leq \deg_T(v)} |V_i|\,|V_j|\,|V_k| &\geq \frac{n-1}{\deg_T(v)}\binom{\deg_T(v)-1}{2} + \binom{\deg_T(v)-1}{3} \\ &= \frac{(\deg_T(v)-3)(n-1)}{2} + \frac{n-1}{\deg_T(v)} + \binom{\deg_T(v)-1}{3} \\ &\geq (n-1)+1+4 = n+4 > n-3.\end{aligned}$$

Thus $\deg_T(v) \leq 4$. In case $\deg_T(v) = 4$, and hence $n \geq 5$, it is easy to compute that the maximum in (4.1) is obtained by $|V_1| = n-4$, $|V_2| = |V_3| = |V_4| = 1$. Therefore we get

$$\sum_{1\leq i<j<k\leq 4} |V_i|\,|V_j|\,|V_k| = (n-4)3+1 = 3n-11,$$

which is greater than $n-3$ for $n \geq 5$.

Altogether we obtain that $\deg_T(v) = 3$ and $T = Q_n$. □

Lemma 4.6. *The tree Q_n maximizes the Wiener index among all connected graphs of order $n \geq 4$ that are different from P_n.*

Proof. Because of Lemma 4.4 we immediately get that

$$W(C_n) < W(Q_n) = \binom{n+1}{3} - (n-3)$$

for $n \geq 4$. Let G be a connected graph of order n different from P_n and C_n. Then G has a spanning tree T that is different from P_n. According to Theorem 3.2 and Lemma 4.5 we obtain

$$W(G) \leq W(T) \leq W(Q_n).$$

□

Furthermore we need some propositions about bipartite graphs:

Lemma 4.7. *Let G be a connected bipartite graph on $a+b$ vertices. Then*

$$W(G) \geq (a+b)(a+b-1) - ab$$

with equality if and only if $G = K_{a,b}$.

Proof. If G is a bipartite graph on $a+b$ vertices different from $K_{a,b}$, it arises from $K_{a,b}$ by deleting some edges. Thus we obtain by Theorem 3.1

$$\begin{aligned} W(G) > W(K_{a,b}) &= ab + 2\sum_{i=1}^{a-1} i + 2\sum_{i=1}^{b-1} i \\ &= ab + a(a-1) + b(b-1) \\ &= (a+b)(a+b-1) - ab. \end{aligned}$$

□

With Lemma 4.7 we immediately obtain the following corollary for $n = a+b$:

Corollary 4.8. *Let G be a connected bipartite graph on n vertices, $n \geq 1$. Then*

$$W(G) \geq \begin{cases} n(n-1) - \frac{n^2}{4} & \text{if } n \text{ is even,} \\ n(n-1) - \frac{n^2-1}{4} & \text{if } n \text{ is odd.} \end{cases}$$

Lemma 4.9. *Let G be a connected bipartite graph on $a+b$ vertices. Then $W(G)$ is odd if and only if both a and b are odd.*

Proof. Let V_1, V_2 be the two vertex sets such that $E(G) = \{(x,y) : x \in V_1, y \in V_2\}$ and $|V_1| = a$, $|V_2| = b$. If $x, y \in V_i$, $i = 1, 2$, it is easy to see that $d_G(x,y) \equiv 0 \mod 2$. Thus we have

$$\frac{1}{2}\sum_{x,y \in V_i} d_G(x,y) \equiv 0 \mod 2.$$

For $x \in V_1$ and $y \in V_2$ we get $d_G(x,y) \equiv 1 \mod 2$ and therefore

$$\sum_{x \in V_1}\sum_{y \in V_2} d_G(x,y) \equiv \begin{cases} 1 \mod 2 & \text{if } ab \equiv 1 \mod 2, \\ 0 \mod 2 & \text{if } ab \equiv 0 \mod 2. \end{cases}$$

Hence $W(G) = \sum_{x \in V_1}\sum_{y \in V_2} d_G(x,y) + \frac{1}{2}\sum_{i=1,2}\sum_{x,y \in V_i} d_G(x,y)$ is odd if and only if $ab \equiv 1 \mod 2$, which again is equivalent to both a and b being odd. □

To find bipartite graphs to solve Problem 4.1 we mainly use four different types of graphs:

Definition 4.3. Let $a \geq 1$ and $1 \leq l \leq k \leq j \leq i \leq a$. Then we denote the bipartite graph with vertex set $\{v_1, \ldots, v_a\} \uplus \{w_1, w_2\}$ and edge set

$$\{(v_b, w_1) : 1 \leq b \leq a\} \cup \{(v_b, w_2) : 1 \leq b \leq i\}$$

by $G_{a,2}(i)$. Furthermore we obtain the bipartite graph $G_{a,3}(i,j)$ on $a+3$ vertices by introducing a new vertex w_3 to $G_{a,2}(i)$ such that the vertex set becomes

$$E(G_{a,2}(i)) \cup \{(v_b, w_3) : 1 \leq b \leq j\}.$$

Analogously the bipartite graph $G_{a,4}(i,j,k)$ on $a+4$ vertices arises from $G_{a,3}(i,j)$ by connecting a new vertex w_4 to it such that the vertex set becomes

$$E(G_{a,3}(i,j)) \cup \{(v_b, w_4) \,:\, 1 \le b \le k\}.$$

Finally we get the bipartite graph $G_{a,5}(i,j,k,l)$ on $a+5$ vertices by introducing a new vertex w_5 to $G_{a,4}(i,j,k)$ such that the vertex set becomes

$$E(G_{a,4}(i,j,k)) \cup \{(v_b, w_5) \,:\, 1 \le b \le l\}.$$

Lemma 4.10. *Let $a \ge 1$ and $1 \le l \le k \le j \le i \le a$. Then*

$$\begin{aligned} W(G_{a,2}(i)) &= a^2 + 3a + 2 - 2i, \\ W(G_{a,3}(i,j)) &= a^2 + 6a + 6 - 2(i+j), \\ W(G_{a,4}(i,j,k)) &= a^2 + 9a + 12 - 2(i+j+k), \\ W(G_{a,5}(i,j,k,l)) &= a^2 + 12a + 20 - 2(i+j+k+l). \end{aligned}$$

Proof. We first show the formula for $W(G_{a,2}(i))$ using the notation of Definition 4.3. The distance between w_1 and v_b, $1 \le b \le a$, is 1 and the distance between w_1 and w_2 is 2. Furthermore it is easy to see that the sum of all distances between $x, y \in \{v_1, \ldots, v_a\}$ is $2\sum_{b=1}^{a-1} b$. At last the distance between w_2 and v_b is one for $1 \le b \le i$ and three for $i+1 \le b \le a$. Therefore we obtain

$$\begin{aligned} W(G_{a,2}(i)) &= a + 2\sum_{b=1}^{a-1} b + i + 3(a-i) + 2 \\ &= a^2 + 3a + 2 - 2i. \end{aligned}$$

Analogously we get

$$\begin{aligned} W(G_{a,3}(i,j)) &= W(G_{a,2}(i)) + j + 3(a-j) + 4 \\ &= a^2 + 6a + 6 - 2(i+j). \end{aligned}$$

The proof of the remaining formulas is similar. □

Notice that the Wiener index of $G_{a,5}(i,j,k,l)$ as well as that of the other bipartite graphs of Definition 4.3 does not depend on the values of i, j, k and l, respectively, but only on their sum. Thus it is obvious that with $a \in \mathbb{N}$ we have the further condition that $2 \le i+j \le 2a$ for $G_{a,3}(i,j)$, $3 \le i+j+k \le 3a$ for $G_{a,4}(i,j,k)$ and $4 \le i+j+k+l \le 4a$ for $G_{a,5}(i,j,k,l)$.

Lemma 4.11. *Let $w, a, i, j, k, l \in \mathbb{N}$ such that $4 \le i+j+k+l \le 4a$. Then there exists a graph $G_{a,5}(i,j,k,l)$ with $W(G_{a,5}(i,j,k,l)) = w$ if and only if*

$$w \in \mathbb{N} \setminus \{0, 1, 2, \ldots, 24, 26, 27, 28, 29, 30, 31, 33, 35, 37, 39, 42, 44, 46, 48, 50, 59, 61, 63, 78\}.$$

Proof. First notice that $W(G_{a,5}(i,j,k,l)) \equiv 1 \mod 2$ if and only if $a \equiv 1 \mod 2$. Furthermore we obtain from Lemma 4.10 that

$$W(G_{a,5}(i,j,k,l)) \in W_a = \{a^2+4a+20, a^2+4a+22, a^2+4a+24, \dots, a^2+12a+12\}$$

where the first value is $W(G_{a,5}(a,a,a,a))$ and the last is $W(G_{a,5}(1,1,1,1))$. Now we try to show that $W_a \cap W_{a+2} \neq \emptyset$ for a big enough. Therefore we compute

$$\begin{aligned} W(G_{a,5}(1,1,1,1)) &- W(G_{a+2,5}(a+2,a+2,a+2,a+2)) \\ &= a^2+12a+12-((a+2)^2+4(a+2)+20) \\ &= 4a-20. \end{aligned}$$

Hence $W(G_{a,5}(1,1,1,1)) \geq W(G_{a+2,5}(a+2,a+2,a+2,a+2))$ for $a \geq 5$. This means that all odd numbers $w_1 \geq 65 = W(G_{5,5}(5,5,5,5))$ and all even numbers $w_0 \geq 80 = W(G_{6,5}(6,6,6,6))$ are Wiener indices of some $G_{a,5}(i,j,k,l)$. Moreover we obtain

$$\begin{aligned} &W(G_{1,5}(i,j,k,l)) \in \{25\}, \\ &W(G_{2,5}(i,j,k,l)) \in \{32,34,36,38,40\}, \\ &W(G_{3,5}(i,j,k,l)) \in \{41,43,45,\dots,57\}, \\ &W(G_{4,5}(i,j,k,l)) \in \{52,54,56,\dots,76\}, \end{aligned}$$

which completes the proof. □

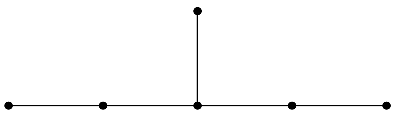
With this we can state the solution to Problem 4.1 for bipartite graphs:

Theorem 4.12. *Let $w \in \mathbb{N}_0$. Then there exists a connected bipartite graph G with Wiener index $W(G) = w$ if and only if $w \in \mathbb{N}_0 \setminus \{2,3,5,6,7,11,12,13,15,17,19,33,37,39\}$.*

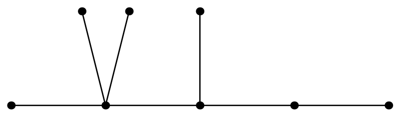
Proof. Direct calculation of $W(G_{a,2}(i))$ with $a = 1,2,3,4,6$ shows that 4, 8, 10, 14, 16, 18, 22, 24, 26, 28, 44, 46,48, 50 and 78 are Wiener indices of connected bipartite graphs. Furthermore by computing $W(G_{a,3}(i,j))$ for $a = 1,3,4$ we obtain that 9, 21, 23, 27, 29, 30 and 42 are Wiener indices of connected bipartite graphs. Since each tree is bipartite and $W(P_1) = 0$, $W(P_2) = 1$, $W(P_5) = 20$ and $W(P_6) = 35$, we get that 0, 1, 20 and 35 are Wiener indices of connected bipartite graphs too. Finally the four bipartite graphs shown in Figure 4.3 are examples of graphs with Wiener index 31, 59, 61 and 63. Thus together with Lemma 4.11 we obtain that there exist bipartite graphs with Wiener index equal to any natural number including zero except those 14 numbers listed above.

Therefore it remains to be shown that none of these 14 numbers can be Wiener indices of bipartite graphs. Because of Corollary 4.8 a bipartite graph G with $W(G) < 8$ must have less than four vertices. The only connected bipartite graphs with less than four vertices are P_1, P_2 and P_3. As their Wiener index is not equal to 2, 3, 5, 6 or 7, these numbers are no Wiener indices of any bipartite graph.

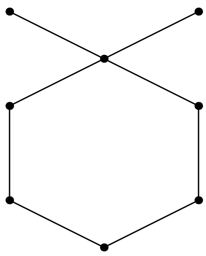
Furthermore Corollary 4.8 implies that the number of vertices of a bipartite graph G with $W(G) < 14$ must be less than five. On the other hand, according to Corollary 3.5 the



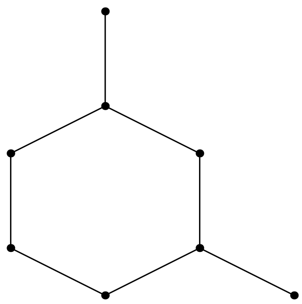
(a) Tree with Wiener index 31.



(b) Tree with Wiener index 63.



(c) Bipartite graph with Wiener index 59.



(d) Bipartite graph with Wiener index 61.

Figure 4.3

maximal Wiener index of a graph with less than five vertices is $W(P_4) = 10$. Thus we get $w \neq 11, 12, 13$.

Again with Corollary 4.8 we obtain that a bipartite graph G with $W(G) < 20$ must have less than six vertices. Furthermore, as Lemma 4.9 implies that the Wiener index of a bipartite graph is even if the number of vertices is odd, we get that a bipartite graph with five vertices can never have Wiener index 15, 17 or 19. But as we already know the maximal Wiener index of a graph with less than five vertices is 10. Therefore 15, 17 and 19 are not Wiener indices of any bipartite graph.

Analogously we obtain by Corollary 4.8 and Lemma 4.9 that a bipartite graph with Wiener index less than 40 must have less than seven vertices. According to Corrollary 3.5 the maximal Wiener index of a graph with less than seven vertices is $W(P_6) = 35$. Hence $w \neq 37, 39$.

Finally, as 33 is odd and less than 40, a bipartite graph with Wiener index 33 must have less than seven vertices. On the other hand, we have $W(P_5) = 20$, which means that the number of vertices must be six. As $W(Q_6) = 32$ and $W(P_6) = 35$, we obtain by Lemma 4.6 that there exists no bipartite graph with Wiener index 33, which completes the proof. $\square$

4.3 Trees

In this section we are going to deal with the question which natural numbers can be Wiener indices of trees. In 1994 Gutman et al. [12], after making some numerical testing, stated the conjecture that there is some bound $M \in \mathbb{N}$ such that for all $w \geq M$ one can find a tree T with $W(T) = w$. Some further checking of all possible trees with 20 and less vertices showed that of all natural numbers up to 1206 there are only 49 numbers which are not

Wiener indices of trees (see [16]). Thus in [16] the conjecture was made that all natural numbers except these 49 numbers are Wiener indices of trees. This was supported by a computational investigation by Ban et al. (see [1]) where they showed that every integer $n \in [10^3, 10^8]$ is the Wiener index of some caterpillar.

Finally, in 2006 Wang and Yu (see [22]) and Wagner (see [20]) independently proved this conjecture. In order to do so Wang and Yu showed that for every integer $w > 10^8$ there exists a short caterpillar with at most six non-leaf vertices such that its Wiener index is w. Wagner, considering another class of trees, was able to give a somehow stronger statement: he could show that every integer ≥ 470 is the Wiener index of some star-like tree.

In the following we will give the proof provided by Wagner.

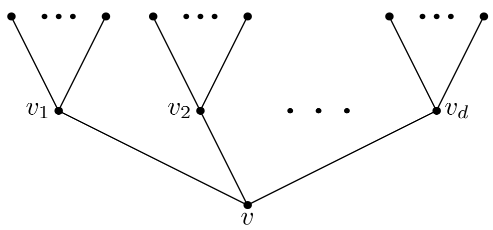


Figure 4.4: The star-like tree $S(c_1, \dots, c_d)$.

Definition 4.4. A tree $S(c_1, \dots, c_d)$ with $n = c_1 + \dots + c_d$ edges is called *star-like* if it arises from the stars $S_{c_1+1}, \dots, S_{c_d+1}$ by taking exactly one leaf of each S_{c_i+1}, $i = 1, \dots, d$, and identifying them with each other. (See Figure 4.4 where the vertex in common is denoted by v and the vertex with degree c_i by v_i, $i = 1, \dots, d$.)

Lemma 4.13. *Let $S(c_1, \dots, c_d)$ be a star-like tree with $n = c_1 + \dots + c_d$. Then*

$$W(S(c_1, \dots, c_d)) = 2n^2 - (d-1)n - \sum_{i=1}^{d} c_i^2.$$

Proof. Let us denote the vertex in the center by v, its neighbours by $v_1, \dots, v_d$ and the leaves by $w_1, \dots, w_{n-d}$. According to the construction of $S(c_1, \dots, c_d)$ the star S_{c_i+1} is a subtree of $S(c_1, \dots, c_d)$, $i = 1, \dots, d$. We already know that $W(S_{c_i+1}) = c_i^2$. Furthermore we have to consider the distance of the following vertex pairs:

- For all $\binom{d}{2}$ pairs v_i, v_j with $1 \leq i < j \leq d$, the distance is 2.
- For v_i, $i = 1, \dots, d$, the sum over its distances to all w_j with $i \neq j$ is $3(n - c_i - d + 1)$.
- The sum over the distances between all pairs of leaves which have no neighbour in common is $4\Big(\binom{n-d}{2} - \sum\limits_{i=1}^{d} \binom{c_i - 1}{2}\Big)$.

Altogether we obtain

$$\begin{aligned}
W(S(c_1,\dots,c_d)) &= \sum_{i=1}^{d} c_i^2 + 2\binom{d}{2} + 3\sum_{i=1}^{d}(n - c_i - d + 1) \\
&\quad + 4\left(\binom{n-d}{2} - \sum_{i=1}^{d}\binom{c_i - 1}{2}\right) \\
&= \sum_{i=1}^{d} c_i^2 + d(d-1) + 3(n-d+1)d - 3n \\
&\quad + 2(n-d)(n-d-1) - 2\sum_{i=1}^{d}(c_i^2 - 3c_i + 2) \\
&= 2n^2 - nd - 5n + 4d - \sum_{i=1}^{d} c_i^2 + 6n - 4d \\
&= 2n^2 - (d-1)n - \sum_{i=1}^{d} c_i^2.
\end{aligned}$$

□

Lemma 4.14. *Let $S(c_1,\dots,c_d)$ be the star-like tree with $n = c_1 + \dots + c_d$. Furthermore let $c_i = c_j = c$ for some $1 \le i, j \le d$, $i \ne j$. Then*

$$W(S(c_1,\dots,c_d)) > W(S(c'_1,\dots,c'_d))$$

for the star-like tree $S(c'_1,\dots,c'_d)$ with

$$c'_k = \begin{cases} c-1 & \text{if } k = i, \\ c+1 & \text{if } k = j, \\ c_k & \text{otherwise.} \end{cases}$$

Proof. Obviously n and d remain unchanged. Thus according to Lemma 4.13 we obtain

$$W(S(c_1,\dots,c_d)) - W(S(c'_1,\dots,c'_d)) = (c-1)^2 + (c+1)^2 - 2c^2 = 2 > 0.$$

□

Definition 4.5. The star-like tree with $c_1 = \dots = c_l = 2$ and $c_{l+1} = \dots = c_{l+k} = 1$ is denoted by $S(2_l, 1_k)$. Besides we refer to the Wiener index of $S(2_l, 1_k)$ as $w(l,k)$.

It is easy to see that for $S(2_l, 1_k)$ we get $n = 2l + k$ and $d = l + k$. Thus we have

$$w(l,k) = 2(2l+k)^2 - (l+k-1)(2l+k) - 4l - k = 6l^2 + (5k-2)l + k^2.$$

l	2	3	4	5	6	7	8	9	10	11	12	13
$s(l)$	1	1	3	4	4	7	9	10	10	14	17	19
l	14	15	16	17	18	19	20	21	22	23	24	25
$s(l)$	20	20	25	29	32	34	35	35	41	46	50	53

Table 4.1: $s(l)$ for all star-like trees $S(2_l, 1_k)$ with $2 \leq l \leq 25$.

Definition 4.6. Let $S(c_1, \ldots, c_d)$ be the star-like tree with $c_i = c_j = c$ for some indices $1 \leq i, j \leq d$, $i \neq j$. Replacing c_i by $c+1$ and c_j by $c-1$ is called a *splitting step.*

The number of splitting steps one can perform starting with $S(2_l, 1_k)$ is denoted by $s(l)$ (which is independent of the number of 1's occurring in $S(2_l, 1_k)$).

Example 4.15. Let us consider the star-like tree $S(2_l, 1_k)$ with $l = 5$. Then we can perform the following splitting steps

$$S(2_5, 1_k) \hookrightarrow S(3, 3, 2, 1_{k+2}) \hookrightarrow S(4, 2, 2, 1_{k+2}) \hookrightarrow S(4, 3, 1_{k+3}).$$

Thus we have $s(5) = 4$.

A list of all $s(l)$ for $2 \leq l \leq 25$ is given in Table 4.1.

Obviously $s(l)$ is a non-decreasing function and we have $s(l+7) \geq s(l) + s(7)$ and $s(l+8) \geq s(l) + s(8)$. Using Table 4.1 we see that $s(l) \geq l + 5$ for $12 \leq l \leq 18$ and $s(l) \geq l + 9$ for $16 \leq l \leq 24$. Thus we obtain $s(l) \geq l + 5$ for $l \geq 12$ and $s(l) \geq l + 9$ for $l \geq 16$.

Now, starting with $S(2_l, 1_k)$, we can construct star-like trees with Wiener index $w(l, k)$, $w(l, k) - 2$, ..., $w(l, k) - 2s(l)$ by using splitting steps and applying Lemma 4.14.

Proposition 4.16. *For every even integer $w \geq 1506$ there exists a star-like tree $S(c_1, \ldots, c_d)$ such that $W(S(c_1, \ldots, c_d)) = w$.*

Proof. Let us consider the star-like tree $S(2_l, 1_k)$ with $k = 0, 2, 4, \ldots, 10$ and $l = x + 1 - \frac{k}{2}$. Then according to Lemma 4.13 we obtain

$$\begin{aligned} W(l, k) &= 6\left(x + 1 - \frac{k}{2}\right)^2 + (5k - 2)\left(x + 1 - \frac{k}{2}\right) + k^2 \\ &= 6x^2 + (10 - k)x + 4. \end{aligned}$$

Let $x \geq 16$. Then we have $l \geq 16 + 1 - \frac{10}{2} = 12$ and therefore

$$s(l) \geq l + 5 \geq l + \frac{k}{2} = x + 1.$$

Thus

$$6x^2 + (10 - k)x + 4 - 2s(l) \leq 6x^2 + (10 - k)x + 4 - 2(x + 1)$$

$$= 6x^2 + (8-k)x + 2$$

and hence all even numbers in the interval

$$[6x^2 + (8-k)x + 2; 6x^2 + (10-k)x + 4]$$

are Wiener indices of star-like trees. Since $6x^2+(10-(k+2))x+4 = [6x^2+(8-k)x+2]+2$ we obtain

$$\bigcup_{k=0,2,\ldots,10} [6x^2 + (8-k)x + 2; 6x^2 + (10-k)x + 4] = [6x^2 - 2x + 2; 6x^2 + 10x + 4].$$

As $6x^2 + 10x + 4 = 6(x+1)^2 - 2(x+1)$, we get

$$\bigcup_{x\geq 16} [6x^2 - 2x + 2; 6x^2 + 10x + 4] = [1506; \infty),$$

which means that all even integers $w \geq 1506$ are Wiener indices of star-like trees. □

Proposition 4.17. *For every odd integer $w \geq 2385$ there exists a star-like tree $S(c_1, \ldots, c_d)$ such that $W(S(c_1, \ldots, c_d)) = w$.*

Proof. We use the same idea as in the proof of Proposition 4.16. So let us consider the star-like trees $S(2_l, 1_k)$ with the pairs $(l,k) = (x-6, 15)$, $(x, 1)$, $(x-4, 11)$, $(x-8, 21)$, $(x-2, 7)$, $(x-6, 17)$ for $x = 2a$, $a \in \mathbb{N}$. Then according to Lemma 4.13 we obtain

$$\begin{aligned} w(x-6,15) &= 6x^2 + x + 3, \\ w(x,1) &= 6x^2 + 3x + 1, \\ w(x-4,11) &= 6x^2 + 5x + 5, \\ w(x-8,21) &= 6x^2 + 7x + 1, \\ w(x-2,7) &= 6x^2 + 9x + 7, \\ w(x-6,17) &= 6x^2 + 11x + 7. \end{aligned}$$

If $x = 2a \geq 20$, we have for all cases that $l \geq 12$. Thus $s(l) \geq l + 5$ and therefore we get that all odd numbers in the intervals

$$\begin{aligned} &[6x^2 - x + 5; 6x^2 + x + 3], \\ &[6x^2 + x - 9; 6x^2 + 3x + 1], \\ &[6x^2 + 3x + 3; 6x^2 + 5x + 5], \\ &[6x^2 + 5x + 7; 6x^2 + 7x + 1], \\ &[6x^2 + 7x + 1; 6x^2 + 9x + 7], \\ &[6x^2 + 9x + 9; 6x^2 + 11x + 7] \end{aligned}$$

are Wiener indices of some star-like trees, which means that all odd numbers in the interval $[6x^2 - x + 5; 6x^2 + 11x + 7]$ for $x = 2a \geq 20$ are Wiener indices of some star-like trees.

Now, let us consider the star-like trees $S(2_l, 1_k)$ with the pairs $(l, k) = (x-1, 3)$, $(x-5, 13)$, $(x-9, 23)$, $(x-3, 9)$, $(x-7, 19)$, $(x-1, 5)$ for $x = 2a-1$, $a \in \mathbb{N}$. Analogously we obtain by Lemma 4.13

$$\begin{aligned}
w(x-1,3) &= 6x^2 + x + 2,\\
w(x-5,13) &= 6x^2 + 3x + 4,\\
w(x-9,23) &= 6x^2 + 5x - 2,\\
w(x-3,9) &= 6x^2 + 7x + 6,\\
w(x-7,19) &= 6x^2 + 9x + 4,\\
w(x-1,5) &= 6x^2 + 11x + 8.
\end{aligned}$$

If $x = 2a - 1 \geq 21$, we have for all cases that $l \geq 12$. Thus $s(l) \geq l + 5$. Furthermore, for $l = x - 3 \geq 18$ we have $s(x-3) \geq (x-3) + 90x + 6$, and therefore we get that all odd numbers in the intervals

$$\begin{aligned}
&[6x^2 - x - 6; 6x^2 + x + 2],\\
&[6x^2 + x + 4; 6x^2 + 3x + 4],\\
&[6x^2 + 3x + 6; 6x^2 + 5x - 2],\\
&[6x^2 + 5x; 6x^2 + 7x + 6],\\
&[6x^2 + 7x + 8; 6x^2 + 9x + 4],\\
&[6x^2 + 9x; 6x^2 + 11x + 8]
\end{aligned}$$

are Wiener indices of some star-like trees, which means that all odd numbers in the interval $[6x^2 - x - 6; 6x^2 + 11x + 8]$ for $x = 2a - 1 \geq 21$ are Wiener indices of some star-like trees.

Combining those two results, we obtain that for any $x \geq 20$ all odd numbers in the interval

$$[6x^2 - x + 4; 6x^2 + 11x + 8]$$

are Wiener indices of some star-like trees. Since $6x^2 + 11x + 8 = 6(x+1)^2 - (x+1) + 3$, we obtain

$$\bigcup_{x \geq 20} [6x^2 - x + 4; 6x^2 + 11x + 8] = [2384; \infty),$$

which means that all odd integer $w \geq 2384$ are Wiener indices of star-like trees. $\square$

Furthermore Wagner checked via computer that all integers $470 \leq w \leq 2384$ are Wiener indices of star-like trees with less then 41 edges. Thus, together with [12], the following theorem holds:

Theorem 4.18. *Let $w \in \mathbb{N}_0$. Then there exists a tree T with $W(T) = w$ if and only if $w \in \mathbb{N}_0 \setminus \{2, 3, 5, 6, 7, 8, 11, 12, 13, 14, 15, 17, 19, 21, 22, 23, 24, 26, 27, 30, 33, 34, 37, 38, 39, 41, 43, 45, 47, 51, 53, 55, 60, 61, 69, 73, 77, 78, 83, 85, 87, 89, 91, 88, 101, 106, 113, 147, 159\}$.*

Bibliography

[1] Y.-E. A. Ban, S. Bereg, and N. H. Mustafa. A conjecture on Wiener indices in combinatorial chemistry. *Algorithmica*, 40(2):99–118, 2004.

[2] D. Bonchev and D. J. Klein. On the Wiener number of thorn trees, stars, rings, and rods. *Croatica Chemica Acta*, 75:613–620, 2002.

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

[4] A. Dobrynin. Branchings in trees and the calculation of the Wiener index of a tree. *MATCH Commun. Math. Comput. Chem.*, 41:119–134, 2000.

[5] A. Dobrynin, R. Entringer, and I. Gutman. Wiener index of trees: theory and application. *Acta Appl. Math.*, 66:211–249, 2001.

[6] A. Dobrynin and I. Gutman. On a graph invariant related to the sum of all distances in a graph. *Publ. Inst. Math. (Beograd)*, 56:18–22, 1994.

[7] J. Doyle and J. Graver. Mean distance in a graph. *Discrete Math.*, 17:147–154, 1977.

[8] R. Entringer, D. Jackson, and D. Snyder. Distance in graphs. *Czechoslovak Math. J.*, 26:283–296, 1976.

[9] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, and L. Volkmann. Wiener index versus maximum degree in trees. *Discrete Applied Mathematics*, 122:127–137, 2002.

[10] I. Gutman. Distance of thorny graphs. *Publ. Inst. Math. (Beograd)*, 63:31–36, 1998.

[11] I. Gutman and Y.-N. Yeh. The sum of all distances in bipartite graphs. *Mathematica Slovaca*, 45:327–334, 1995.

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

[13] M. Hazewinkel, editor. *Encyclopaedia of mathematics*. Kluwer Academic Publishers, 2002. An online version can be found at `http://eom.springer.de/default.htm`.

[14] N. Jacobson. *Basic algebra I.* W. H. Freeman and company (San Francisco), 1974.

[15] W. Kocay and D. L. Kreher. *Graphs, algorithms, and optimization.* Chapman and Hall / CRC, 2005.

[16] M. Lepović and I. Gutman. A collective property of trees and chemical trees. *J. Chem. Inf. Comput. Sci.*, 38:823–826, 1998.

[17] R. Merris. An edge version of the matrix-tree theorem and the Wiener index. *Linear and Multilinear Algebra*, 25:291–296, 1989.

[18] O. E. Polansky and D. Bonchev. The Wiener number of graphs. I. General theory and changes due to some graph operations. *MATCH Commun. Math. Comput. Chem.*, 21:133–186, 1986.

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

[20] S. G. Wagner. A class of trees and its Wiener index. *Acta Appl. Math.*, 91:119–132, 2006.

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

[22] H. Wang and G. Yu. All but 49 numbers are Wiener indices of trees. *Acta Appl. Math.*, 92:15–20, 2006.

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

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

[25] X.-D. Zhang, Y. Liu, and M.-X. Han. The maximum Wiener index of trees with given degree sequences. Preprint, arXiv:0907.3772v1, July 2009.