

# *Hyper-Wiener Index vs. Wiener Index.* **Two Highly Correlated Structure-Descriptors**

Ivan Gutman\* and Boris Furtula

Faculty of Science, University of Kragujevac, P.O. Box 60,  
YU-34000 Kragujevac, Yugoslavia

Received December 10, 2002; accepted December 17, 2002  
Published online May 26, 2003 © Springer-Verlag 2003

**Summary.** The *Wiener* ( $W$ ) and *hyper-Wiener* ( $WW$ ) indices of alkanes are found to be highly correlated. Hence, these two structure-descriptors pertain to the very same structural features of the underlying molecules and one of them may be viewed as superfluous. For alkane isomers with  $n$  carbon atoms,  $WW$  is bounded from both above and below by linear functions of  $W$ . The upper bound  $(n/4 + 2)W - n(n-1)(n+1)/4$  and the lower bound  $(3/2)W - (n-1)/2$  for  $W \leq W_0$  and  $(3n/4)W - n(n-1)^2(n+1)$  for  $W \geq W_0$ , where  $W_0 = (2/3)(n-1)(n^3 - n - 1)/(n-2)$  are better than the previously reported estimates of the same kind. In spite of this, the correlation between  $W$  and  $WW$  is curvilinear.

**Keywords.** *Hyper-Wiener* index; *Wiener* index; Alkanes; Chemical graph theory.

## Introduction

The *Wiener index* [1] ( $W$ ) – the sum of distances between all pairs of vertices of the molecular graph – belongs among the oldest graph-based structure descriptors (= “topological indices”) [2–4]. Numerous of its chemical applications were reported [2, 5–7] and its mathematical properties are understood in due detail [2, 8, 9].

Ten years ago another graph-based structure descriptor was introduced [10], named *hyper-Wiener index* and denoted by  $WW$ . It is defined as follows: Let  $T$  be a tree and let  $u$  and  $v$  be two of its vertices, connected by a (unique) path  $\pi_{uv}$ . Let  $n_1(u, v)$  and  $n_2(u, v)$  denote the number of vertices of  $T$ , lying on the two sides of  $\pi_{uv}$ . Then  $WW$  is given by Eq. (1) with the summation going over all pairs of vertices (both adjacent and non-adjacent) of  $T$ .

$$WW = WW(T) = \sum_{u,v} n_1(u, v) \cdot n_2(u, v) \quad (1)$$

---

\* Corresponding author. E-mail: gutman@knez.uis.kg.ac.yu

Note that if the summation on the right-hand side of Eq. (1) is restricted to adjacent vertices  $u, v$ , then it becomes equal to the ordinary *Wiener* index  $W = W(T)$ .

The *hyper-Wiener-index* concept rapidly gained popularity among mathematical chemists [4] and its applications in QSPR and QSAR studies as well as various methods for its computation followed [11–26]. Curiously, however, the relation between the *hyper-Wiener* and the *Wiener* indices was not investigated until quite recently [27].

In Ref. [27] it was found that for a fixed value  $n$  of the number of vertices of the molecular graph,  $WW$  is bounded from both below and above by linear functions of  $W$ . This observation suggests that a linear correlation between  $WW$  and  $W$  might exist. In this work we examine the connection between  $WW$  and  $W$  more carefully and find that, indeed, these structure-descriptors are highly correlated but the correlation between them is not linear. Further, we now significantly improve the earlier reported [27] bounds for  $WW$ .

## Results and Discussion

### *Estimating WW*

A tree is a connected acyclic graph [2]. A tree in which no vertex has a degree greater than 4 is called a chemical tree. Chemical trees are the molecular graphs of alkanes [2].

In Ref. [27] the inequalities shown in Eqs. (2) and (3) were deduced, valid for all  $n$ -vertex trees  $T$  (and thus valid for the molecular graphs of all isomeric alkanes with  $n$  carbon atoms).

$$WW(T) \geq \frac{3n}{4}W(T) - \frac{1}{12}n(n-1)^2(n+1) \quad (2)$$

$$WW(T) \leq \left(\frac{n}{2} + 1\right)W(T) - \frac{1}{2}(n^2 - 2n + 2)(n-1). \quad (3)$$

Note that in Ref. [27], the multiplier  $(n/2 + 1)$  in Eq. (3) was erroneously given as  $(n/2 - 1)$ .

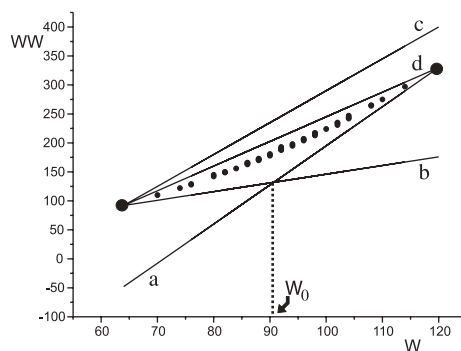
There exist somewhat better estimates of the same kind. In the Appendix we show that Eqs. (2) and (3) are improved by Eqs. (4) and (5), respectively.

$$WW(T) \geq \frac{3}{2}W(T) - \frac{1}{2}(n-1) \quad (4)$$

$$WW(T) \leq \left(\frac{n}{4} + 2\right)W(T) - \frac{1}{4}n(n-1)(n+1), \quad (5)$$

More precisely, the lower bound given in Eq. (4) is better than that in Eq. (2) only for  $W \leq W_0$ , where  $W_0 = (2/3)(n-1)(n^3 - n - 1)/(n-2)$  is the solution of Eq. (6).

$$\frac{3n}{4}W(T) - \frac{1}{12}n(n-1)^2(n+1) = \frac{3}{2}W(T) - \frac{1}{2}(n-1) \quad (6)$$



**Fig. 1.** *Hyper-Wiener* indices ( $WW$ ) of trees with  $n = 9$  vertices versus the respective *Wiener* indices ( $W$ ); heavy dots indicate the star (smallest  $W$  and  $WW$ ) and the path (greatest  $W$  and  $WW$ ); lines  $a$  and  $b$  pertain, respectively, to the lower bounds given by Eqs. (2) and (4); lines  $c$  and  $d$  stand for the upper bounds given by Eqs. (3) and (5), respectively; combination of  $b$  (for  $W \leq W_0$ ) and  $a$  (for  $W \geq W_0$ ), Eq. (6), provides a still better lower bound for the *hyper-Wiener* index

In view of this, the best lower bound for the *hyper-Wiener* index that we can offer is given in Eq. (7).

$$WW(T) \geq \begin{cases} \frac{3}{2}W(T) - \frac{1}{2}(n-1) & \text{for } W \leq W_0 \\ \frac{3n}{4}W(T) - \frac{1}{12}n(n-1)^2(n+1) & \text{for } W \geq W_0. \end{cases} \quad (7)$$

The quality of the estimates of Eqs. (2)–(6) is illustrated in Fig. 1.

### *Correlation between Wiener and Hyper-Wiener Index*

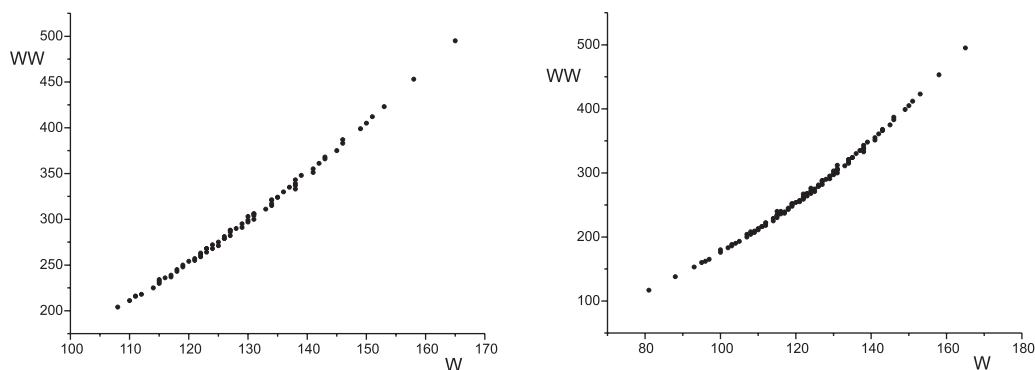
The existence of lower and upper bounds for the *hyper-Wiener* index, that both are linear functions of the *Wiener* index, suggests that there must exist some correlation between  $W$  and  $WW$  and that it might also be linear [27]. In order to examine the validity of this hypothesis, we have undertaken extensive numerical testing. Our results only partially corroborate the earlier expectations.

Our main findings are the following: (1) Within sets of chemical trees (*i.e.*, within sets of isomeric alkanes), as well as within sets of general trees with a fixed number  $n$  of vertices, there is a remarkably good correlation between  $W$  and  $WW$ . (2) This correlation is curvilinear.

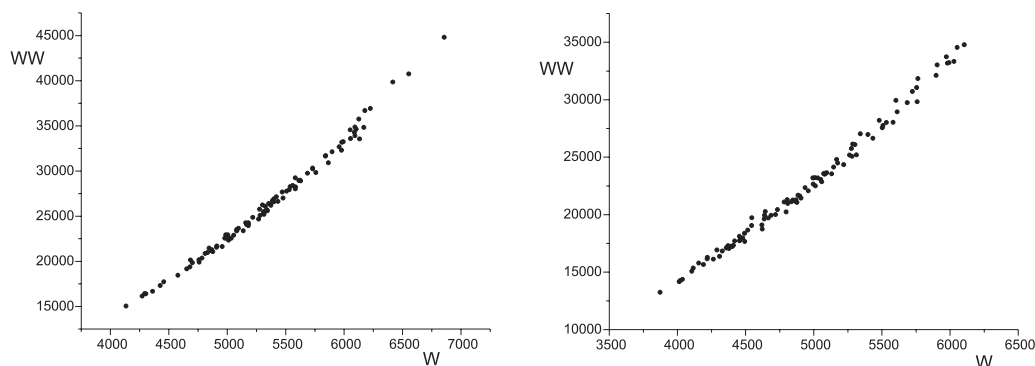
In Fig. 2 are shown two typical examples, the  $WW/W$ -correlation for chemical and general trees with  $n = 10$ . Here, all possible trees have been considered. In Fig. 3 are shown analogous examples for  $n = 40$ . Because the number of alkane isomers and general trees with  $n = 40$  is enormously large, we have used samples consisting of 100 chemical or general trees, constructed uniformly by random using an appropriate random-tree generator [28].

Some characteristic data on the correlations examined are collected in Table 1.

In view of the fact that both  $W$  and  $WW$  are easily computed [11–15, 18, 23–25], we did not intend to design a formula that would enable the (approximate) evaluation of  $WW$  from the known  $W$ -value. Therefore, we examined only



**Fig. 2.** Correlation between the *Wiener* and *hyper-Wiener* indices for the 75 isomeric decanes (left) and for the 106 ten-vertex trees (right); for details see Table 1



**Fig. 3.** Correlation between the *Wiener* and *hyper-Wiener* indices for 100 randomly selected 40-vertex chemical trees (left) and for 100 randomly selected 40-vertex general trees (right); for details see Table 1

the linear and quadratic functional dependence between  $W$  and  $WW$ , namely  $WW \approx a_1W + a_0$  and  $WW \approx b_2W^2 + b_1W + b_0$ , aiming at proving (or disproving) the non-linearity of the respective correlation. The data shown in Table 1 convincingly confirm that the correlation is remarkably good and that it is non-linear. (Recall that we don't claim that it is quadratic, just that the quadratic probe function gives significantly better results than the linear one. This suffices to infer curvilinearity.)

Next, what can be observed from the correlation coefficients given in Table 1 is that there is hardly any difference between the behavior of chemical trees and general trees. Thus, the close analogy in the structure-dependence of  $W$  and  $WW$  is a generally valid property, by no means limited to chemical trees.

In all cases examined the multiplier  $b_2$  (computed by least-squares fitting) was found to be positive-valued, implying that the function describing the correlation between  $W$  and  $WW$  is convex.

**Table 1.** The correlation coefficients for the linear ( $R_{\text{lin}}$ ) and quadratic ( $R_{\text{quad}}$ ) correlation between the *Wiener* and the *hyper-Wiener* indices of chemical trees (CT) and general trees (GT) possessing  $n$  vertices; note that in all cases  $R_{\text{quad}}$  is significantly closer to unity than  $R_{\text{lin}}$ ; #CT and #GT are the respective sample sizes; up to  $n = 10$  all possible chemical and general trees were considered, whereas for  $n \geq 20$  the samples examined consisted of 100 randomly generated species; the curvilinearity of all the examined correlations was found to be statistically significant (by means of  $F$ -test at a 99% confidence level); exceptionally, in the cases  $n = 6$  and  $n = 7$ , for chemical trees the curvilinearity could be verified only at a 95% confidence level; in all examined cases the curve describing the correlation is convex

$n$	CT			GT		
	#CT	$R_{\text{lin}}$	$R_{\text{quad}}$	#GT	$R_{\text{lin}}$	$R_{\text{quad}}$
6	5	0.9987	0.9998	6	0.9941	0.9999
7	9	0.9955	0.9993	11	0.9905	0.9996
8	18	0.9944	0.9993	23	0.9888	0.9994
9	35	0.9945	0.9990	47	0.9869	0.9991
10	75	0.9937	0.9987	106	0.9861	0.9988
20	100	0.9926	0.9985	100	0.9870	0.9960
30	100	0.9903	0.9976	100	0.9845	0.9975
40	100	0.9874	0.9955	100	0.9894	0.9949

## Conclusion

It is somewhat surprising that the existence of the excellent correlation between the *Wiener* index and its 40 years younger congener, the *hyper-Wiener* index, was not earlier reported. One of the first steps to be done after a new topological index is put forward, should be the testing of its independence of the already known indices, especially of those that are conceptually so closely related as  $WW$  and  $W$ .

Anyway, we have now shown that both  $W$  and  $WW$  have the very same dependency on the structure of the underlying molecule, implying that their simultaneous usage in QSPR and QSAR studies would be redundant. We have demonstrated this for chemical trees, *i.e.*, for alkanes. It may be that in the case of other types of compounds (especially those other than hydrocarbons, whose molecular graphs possess weighted edges and vertices) there is a greater degree of dissimilarity between the behavior of  $W$  and  $WW$ . If so, then the *hyper-Wiener* index would deserve to be retained among the molecular structure-descriptors that are attractive and useful from a practitioner's point of view. This (unlikely) possibility would first have to be convincingly demonstrated, which the present authors are not intending to attempt.

## Appendix

### *A Lower Bound for WW in Terms of W*

For a tree  $T$  the relation shown in Eq. (8) between  $WW$  and  $W$  is known [15] where  $d(u, v)$  is the distance between the vertices  $u$  and  $v$ , and where the summation goes

over all pairs of vertices of  $T$ .

$$WW(T) = \frac{1}{2}W(T) + \frac{1}{2}\sum_{u,v} d(u,v)^2 \quad (8)$$

Recall that in this notation  $W(T)$  is given by Eq. (9).

$$W(T) = \sum_{u,v} d(u,v) \quad (9)$$

Now, the sum on the right-hand side of Eq. (8) can be rewritten as shown in Eq. (10) with  $\sum^{\text{adj}}$  and  $\sum^{\text{nonadj}}$  indicating summation over pairs of adjacent and non-adjacent vertices, respectively.

$$\sum_{u,v} d(u,v)^2 = \sum_{u,v}^{\text{adj}} d(u,v)^2 + \sum_{u,v}^{\text{nonadj}} d(u,v)^2 \quad (10)$$

An  $n$ -vertex tree has  $n-1$  edges, *i.e.*,  $n-1$  pairs of adjacent vertices. For adjacent vertices it is  $d(u,v) = 1$  and thus, Eq. (11) is true.

$$\sum_{u,v}^{\text{adj}} d(u,v)^2 = n - 1. \quad (11)$$

For non-adjacent vertices  $d(u,v) \geq 2$ , and therefore  $d(u,v)^2 \geq 2d(u,v)$ , implying Eq. (12).

$$\begin{aligned} \sum_{u,v}^{\text{nonadj}} d(u,v)^2 &\geq 2 \sum_{u,v}^{\text{nonadj}} d(u,v) = 2 \left[ \sum_{u,v} d(u,v) - (n-1) \right] \\ &= 2[W(T) - (n-1)] \end{aligned} \quad (12)$$

Substituting Eqs. (11) and (12) back into Eq. (10) yields Eq. (13) which combined with Eq. (8) results in the lower bound defined by Eq. (4).

$$\sum_{u,v} d(u,v)^2 \geq 2W(T) - (n-1) \quad (13)$$

Note that equality in Eq. (4) occurs if all pairs of non-adjacent vertices are at distance 2, *i.e.*, in the case of stars.

#### *An Upper Bound for $WW$ in Terms of $W$*

The statistical analysis outlined in a previous section clearly shows that the correlation between the *Wiener* and the *hyper-Wiener* index is curvilinear and convex. Any convex function  $f(x)$  has the following property: Let  $x_1 < x_2$ , and let  $g(x)$  be the straight line drawn through the points  $(x_1, f(x_1))$  and  $(x_2, f(x_2))$ . Then  $g(x) > f(x)$  holds for any  $x$ ,  $x_1 < x < x_2$ , *i.e.*, in the interval  $[x_1, x_2]$ ,  $g(x)$  is an upper bound for  $f(x)$ .

The fact that among all  $n$ -vertex trees the star  $S_n$  has minimum  $W$ - and  $WW$ -values, and the path  $P_n$  has maximum  $W$ - and  $WW$ -values is long known [2, 18, 19].

Further, Eq. (14) is valid.

$$\begin{aligned} W(S_n) &= (n-1)^2; & WW(S_n) &= \frac{1}{2}(n-1)(3n-4) \\ W(P_n) &= \binom{n+1}{3}; & WW(P_n) &= \binom{n+2}{4} \end{aligned} \quad (14)$$

The upper bound given by Eq. (5) is just the straight line drawn through the points  $(W(S_n), WW(S_n))$  and  $(W(P_n), WW(P_n))$ .

## References

- [1] Wiener H (1947) *J Am Chem Soc* **69**: 17
- [2] Gutman I, Polansky OE (1986) *Mathematical Concepts in Organic Chemistry*. Springer, Berlin Heidelberg New York Tokyo
- [3] Mihalić Z, Trinajstić N (1992) *J Chem Educ* **69**: 701
- [4] Todeschini R, Consonni V (2000) *Handbook of Molecular Descriptors*. Wiley-VCH, Weinheim
- [5] Needham DE, Wei IC, Seybold PG (1988) *J Am Chem Soc* **110**: 4186
- [6] Gutman I, Yeh YN, Lee SL, Luo YL (1993) *Indian J Chem* **32A**: 651
- [7] Rücker G, Rücker C (1999) *J Chem Inf Comput Sci* **39**: 788
- [8] Dobrynin AA, Entringer R, Gutman I (2001) *Acta Appl Math* **66**: 211
- [9] Dobrynin AA, Gutman I, Klavžar S, Žigert P (2002) *Acta Appl Math* **72**: 247
- [10] Randić M (1993) *Chem Phys Lett* **211**: 478
- [11] Lukovits I, Linert W (1994) *J Chem Inf Comput Sci* **34**: 899
- [12] Lukovits I (1994) *J Chem Inf Comput Sci* **34**: 1079
- [13] Lukovits I (1995) *Comput Chem* **19**: 27
- [14] Linert W, Renz F, Kleestorfer K, Lukovits I (1995) *Comput Chem* **19**: 395
- [15] Linert W, Kleestorfer K, Renz F, Lukovits I (1995) *J Mol Struct (Theochem)* **337**: 121
- [16] Klein DJ, Lukovits I, Gutman I (1995) *J Chem Inf Comput Sci* **35**: 50
- [17] Diudea MV (1996) *J Chem Inf Comput Sci* **36**: 833
- [18] Linert W, Lukovits I (1997) *MATCH – Commun Math Comput Chem* **35**: 65
- [19] Gutman I, Linert W, Lukovits I, Dobrynin AA (1997) *J Chem Inf Comput Sci* **37**: 349
- [20] Gutman I (1997) *Indian J Chem* **36A**: 128
- [21] Dobrynin AA, Gutman I, Piottukh-Peletsii VN (1999) *J Struct Chem* **40**: 293
- [22] Plavšić D, Lerš N, Sertić-Bionda K (2000) *J Chem Inf Comput Sci* **40**: 516
- [23] Žigert P, Klavžar S, Gutman I (2000) *ACH Models Chem* **137**: 83
- [24] Klavžar S, Žigert P, Gutman I (2000) *Comput Chem* **24**: 229
- [25] Aringhieri R, Hansen P, Malucelli F (2001) *J Chem Inf Comput Sci* **41**: 958
- [26] Lukovits I, In: Diudea MV (Ed) (2001) *QSPR/QSAR Studies by Molecular Descriptors*. Nova, Huntington, pp 31–38
- [27] Gutman I (2002) *Chem Phys Lett* **364**: 352
- [28] Wilf H (1981) *J Algorithms* **2**: 204