

## Three–Center Harary Index and its Applications

**BORIS FURTULA, IVAN GUTMAN, VLADIMIR KATANIĆ**

Faculty of Science, University of Kragujevac, P. O. Box 60, 34000 Kragujevac, Serbia

Correspondence should be addressed to gutman@kg.ac.rs (Ivan Gutman)

Received 18 November 2015; Accepted 6 January 2016

ACADEMIC EDITOR: ALI REZA ASHRAFI

**ABSTRACT** The Harary index  $H$  can be viewed as a molecular structure descriptor composed of increments representing interactions between pairs of atoms, such that their magnitude decreases with the increasing distance between the respective two atoms. A generalization of the Harary index, denoted by  $H_k$ , is achieved by employing the Steiner-type distance between  $k$ -tuples of atoms. We show that the linear combination  $H + \lambda H_3$  is significantly better correlated with a variety of physico-chemical properties of alkanes than  $H$  itself.

**KEYWORDS** Harary index • multicenter Harary index • Steiner distance • molecular graph.

### 1. INTRODUCTION

Let  $G$  be a molecular graph [1] and  $v_1, v_2, \dots, v_n$  be its vertices. The distance [1,2] between the vertices  $v_i$  and  $v_j$ , denoted by  $d(v_i, v_j)$ , is the number of edges in (= the length of) a shortest path that connects  $v_i$  and  $v_j$ . The oldest distance-based molecular structure descriptor is the Wiener index, defined as [3,4]

$$W = W(G) = \sum_{i < j} d(v_i, v_j). \quad (1)$$

Although this topological index found numerous chemical applications, a general objection to it is that pairs of vertices at the greatest distance have the greatest contributions to the numerical value of  $W$ . Bearing in mind that vertices of a molecular graph represent atoms [1], this property of the Wiener index seemingly contradicts the fact that the interaction between near-lying atoms are greatest. Several attempts were made to eliminate this difficulty [5-7], but the simplest and more efficient solution was achieved by modifying Eq. (1) as

$$H = H(G) = \sum_{i < j} \frac{1}{d(v_i, v_j)}. \quad (2)$$

Formula (2) was independently conceived by Ivanciuc et al. [8] and Plavšić et al. [9]. The molecular structure descriptor  $H$  was named “*Harary index*” (in honor of Frank Harary [10]). Eventually, this index attracted much attention; see the surveys [11,12], the recent book [13], and the recent papers [14–16].

Same as the Wiener index, Eq. (1), the Harary index  $H$ , Eq. (2), may also be viewed as a sum of structural increments representing *pairs* of atoms, i.e., *two-center* interatomic interactions. Following an idea outlined in a recent paper [17], one could think of three-center, four-center, etc. interactions that would lead to the following evident multicenter extension of the Harary-index concept:

$$H_3 = H_3(G) = \sum_{i < j < k} \frac{1}{d(v_i, v_j, v_k)} \quad (3)$$

$$H_4 = H_4(G) = \sum_{i < j < k < l} \frac{1}{d(v_i, v_j, v_k, v_l)}$$

and in the general case,

$$H_k = H_k(G) = \sum_{i_1 < i_2 < \dots < i_k} \frac{1}{d(v_{i_1}, v_{i_2}, \dots, v_{i_k})}.$$

In the above formulas, for the multiple-vertex distances we take the standard *Steiner distance* (introduced in graph theory as early as in the 1980s [18]). For details on the Steiner distance see the recent paper [17] or in appropriate monographs [19-21].

In nutshell: The Steiner distance  $d(v_{i_1}, v_{i_2}, \dots, v_{i_k})$  of  $k$  distinct vertices  $v_{i_1}, v_{i_2}, \dots, v_{i_k}$  of a graph  $G$ , is the number of edges of a connected subgraph of  $G$ , with smallest number of edges, containing all the vertices  $v_{i_1}, v_{i_2}, \dots, v_{i_k}$ .

The multicenter Harary indices, based on Steiner distances of  $k$ -tuples of vertices, will be denoted as  $H_k$ . For reasons explained in the subsequent section, in what follows we shall be concerned only with the three-center Harary index  $H_3$ .

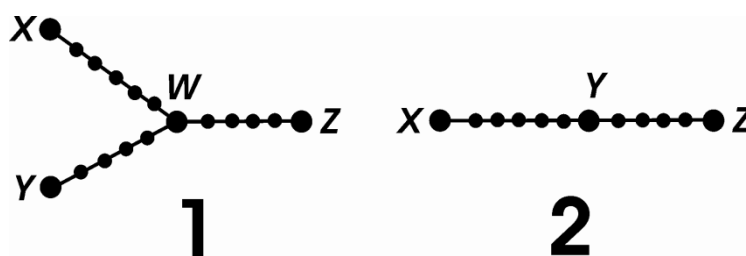
## 2. AN AUXILIARY LEMMA

In the general case, the calculation of the Steiner distance  $d(v_{i_1}, v_{i_2}, \dots, v_{i_k})$  of a  $k$ -tuple of vertices of a molecular graph is a difficult and computation-extensive task. In the case  $k = 3$ , this calculation is significantly simplified by means of the following Lemma:

**Lemma 1.** Let  $G$  be a (connected) graph and  $x, y, z$  its three distinct vertices. Then the Steiner distance of  $x, y, z$  is related to the ordinary distance of these vertices as:

$$d(x, y, z) = \frac{1}{2} [d(x, y) + d(x, z) + d(y, z)]. \quad (4)$$

*Proof.* Any connected subgraph of  $G$  with the smallest number of edges is necessarily a tree (i.e., it is acyclic). We have to distinguish between two cases: when the three vertices  $x, y, z$  are not collinear and when these are collinear, see Figure 1.



**Figure 1.** Two different arrangements of the vertices  $x, y, z$  in a molecular graph; for details see the proof of Lemma 1.

**Case 1:**  $x, y, z$  are not collinear. Then, using the notation specified in Fig. 1,

$$d(x, y, z) = d(x, w) + d(y, w) + d(z, w). \quad (5)$$

On the other hand,

$$d(x, y) = d(x, w) + d(y, w)$$

$$d(x, z) = d(x, w) + d(z, w)$$

$$d(y, z) = d(y, w) + d(z, w)$$

which yields  $d(x, y) + d(x, z) + d(y, z) = 2[d(x, w) + d(y, w) + d(z, w)]$ . Then Eq. (4) follows by bearing in mind (5).

**Case 2:**  $x, y, z$  are collinear. Then, using the notation specified in Fig. 1,

$$d(x, y, z) = d(x, z). \quad (6)$$

Then we have

$$d(x, z) = \frac{1}{2}[d(x, z) + d(x, z)] = \frac{1}{2}[d(x, y) + d(y, z) + d(x, z)]$$

and Eq. (4) directly follows from (6).

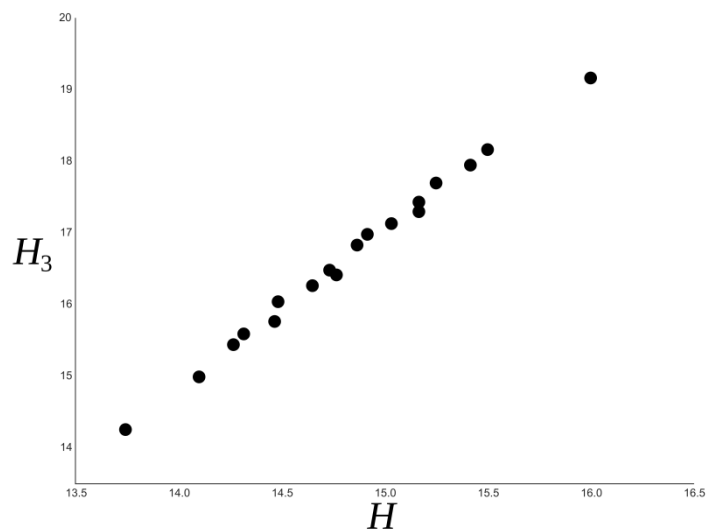
This completes the proof of Lemma 1.

By means of Lemma 1, the calculation of the three-center Harary index, Eq. (3), becomes quite easy and is of equal (low) complexity as the calculation of the respective Wiener index. Unfortunately, results analogous to Lemma 1 could not be established for  $k \geq 4$ .

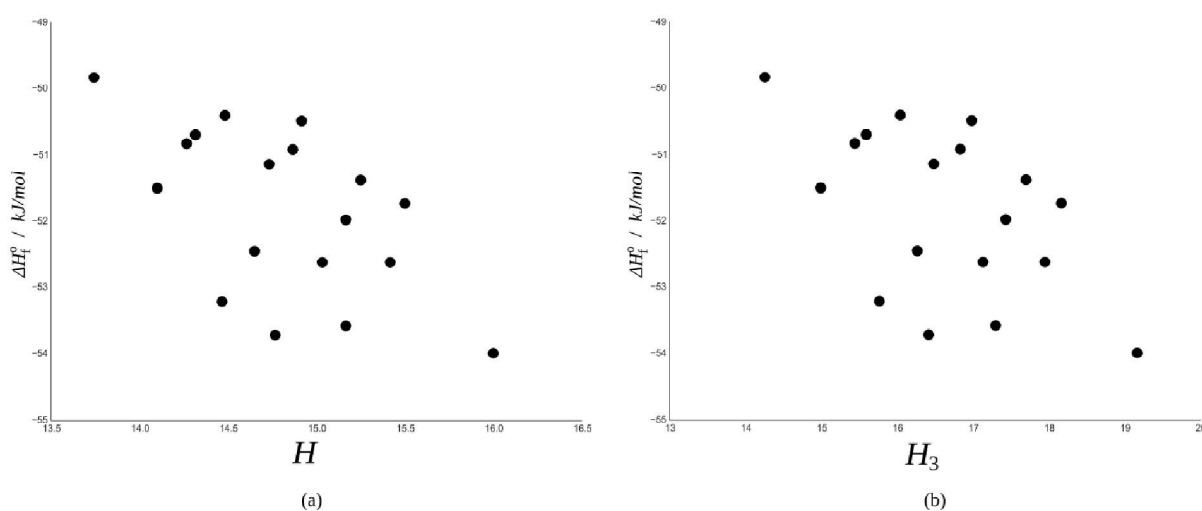
### 3. NUMERICAL WORK

We first present two results that look rather discouraging. In Figure 2 is shown the correlation between the Harary and three-center Harary indices in the case of isomeric octanes. In Figure 3a is shown the correlation between Harary index and standard enthalpy of formation ( $\Delta H_f^\circ$ ) of the same set of octanes [22]. In Figure 3b is displayed the analogous plot for the three-center Harary index. In view of the very good linear

correlation between  $H$  and  $H_3$  (cf. Figure 2), there is no statistical difference between the two correlations shown in Figure 3, and both are disappointingly weak.

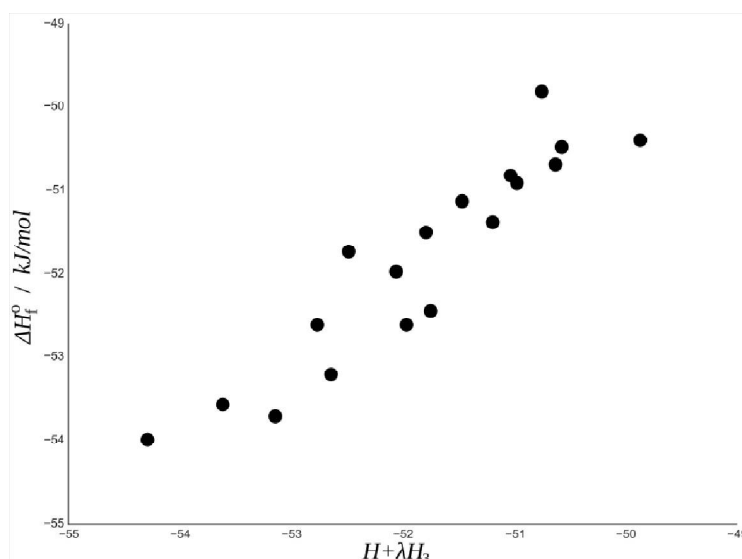


**Figure 2.** Correlation between the three-center Harary index  $H_3$ , Eq. (3), and the ordinary Harary index  $H$ , Eq. (2) for the set of isomeric octanes (18 data points); the correlation coefficient is  $R = 0.9980$ .



**Figure 3.** (a) Correlation between the standard enthalpy of formation ( $\Delta H_f^\circ$ ) of isomeric octanes [22] and the ordinary Harary index  $H$ ;  $R = -0.576$ .  
(b) Analogous correlation with the three-center Harary index  $H_3$ ;  $R = -0.528$ .

A remarkable improvement is obtained by means of the linear combination of the two Harary indices, namely  $H + \lambda H_3$ , shown in Figure 4, where the optimized value  $\lambda = -0.443$  is used for the parameter  $\lambda$ .



**Figure 4.** Correlation between the standard enthalpy of formation ( $\Delta H_f^o$ ) of isomeric octanes [22] and the linear combination  $H + \lambda H_3$  of the two Harary indices. The best results are obtained for  $\lambda = -0.433$  in which case the correlation coefficient increases to  $R = 0.928$ .

Analogous improvements have been found also in the case of a number of other physico-chemical properties of octanes. The respective statistical data are collected in Table 1.

Property	$R(H)$	$R(H_3)$	$R(H + \lambda H_3)$	$\lambda$
$\Delta H_f^o$	-0.576	-0.542	0.928	-0.433
$S^o$	-0.929	-0.914	0.954	-0.356
$\Delta H_{ev}$	-0.779	-0.745	0.928	-0.414
$BP$	-0.573	-0.533	0.831	-0.429
$CT$	-0.111	-0.063	0.756	-0.451
$CP$	0.505	0.540	0.754	-0.483
$\log P$	-0.184	-0.192	0.223	-0.503

**Table 1.** Correlation coefficients for the correlation between physico-chemical properties of octane isomes [22] and Harary index ( $R(H)$ ), three-center Harary index ( $R(H_3)$ ), and the linear combination thereof ( $R(H + \lambda H_3)$ );  $\Delta H_f^o$  = standard enthalpy of formation,  $S^o$  = standard entropy,  $\Delta H_{ev}$  = enthalpy of evaporation,  $BP$  = boiling point at atmospheric pressure,  $CT$  = critical temperature,  $CP$  = critical pressure,  $\log P$  = logarithm of octanol/water partition coefficient; in the last column is the optimized value of the parameter  $\lambda$  for which  $R(H + \lambda H_3)$  is maximal.

#### 4. DISCUSSION AND CONCLUDING REMARKS

If the topological indices  $H$  and  $H_3$  were exactly linearly correlated, then their linear combination would not result in any improvement. From the data shown in Table 1 we see that in some cases significant improvements are obtained, which may be viewed as a kind of (convenient) surprise. This especially is the case for formation enthalpy, critical temperature, and critical pressure. Remarkably, whereas the indices  $H$  and  $H_3$  are completely uncorrelated with critical temperature, i.e.,  $R(H) \approx R(H_3) \approx 0$ , their linear combination results in a moderately good correlation,  $R(H + \lambda H_3) > 0.75$ .

In all cases, the linear combination of  $H$  and  $H_3$  improves the quality of the correlations. However, in some cases, the gain is minor and insignificant. This (necessarily) happens for properties that are well correlated with  $H$  and  $H_3$  (e.g., entropy), but also when the initial correlations are weak (e.g. partition coefficient).

A noteworthy fact is that the optimized value of the parameter  $\lambda$  is nearly equal for all physico-chemical properties examined, and is always negative-valued.

The Harary index is constructed so as to take into account (in a rather rough manner) interactions between pairs of atoms. In the case of usually employed molecular graphs, these are pairs of carbon atoms. At the first glance, by means of the three-center version of the Harary index, some more subtle interatomic interactions might be also taken into account. However, the fact that the parameter  $\lambda$  is relatively large and always negative, indicates that the ordinary Harary index seems to overestimate the interactions between pairs of atoms. Thus, the principal role of  $H_3$  would be to provide a compensation for the intrinsic error of the Harary index. As we could see, this compensation is often very efficient.

From a practitioner's point of view, neither the original Harary index, nor its version improved by the three-center index, are sufficient to accurately reproduce a physico-chemical (or any other) property of the considered organic compounds. This is clearly seen by inspecting Figures 2 and 3. This apparent deficiency of graph-based structure descriptors is nowadays well understood and individual descriptors are nowadays hardly ever used for modeling properties of organic molecules. The successful strategy is to simultaneously use several such descriptors and construct QSPR/QSAR models [4,23–25]. The interested reader is referred to the recent survey [25] in which the design of several commercially interesting pharmacologically active substances (including those with anti-cancer activity) is described.

The present studies indicate that the efficiency of QSPR/QSAR models would be much improved if instead of the Harary index, its combination with three-center Harary index would be utilized.

#### REFERENCES

- 1 I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986.
- 2 F. Buckley, F. Harary, *Distance in Graphs*, Addison-Wesley, Redwood, 1990.

- 3 H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.
- 4 J. Devillers, A. T. Balaban, Eds., *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon & Breach, Amsterdam, 1999.
- 5 M. V. Diudea, O. Minailiuc, A. T. Balaban, Molecular topology. IV. Regressive vertex degrees (new graph invariants) and derived topological indices, *J. Comput. Chem.* **12** (1991) 527–535.
- 6 Z. Mihalić, N. Trinajstić, A graph-theoretical approach to structure-property relationships, *J. Chem. Educ.* **69** (1992) 701–712.
- 7 A. T. Balaban, M. V. Diudea, Real number vertex invariants: regressive distance sums and related topological indices, *J. Chem. Inf. Comput. Sci.* **33** (1993) 421–428.
- 8 O. Ivanciuc, T. S. Balaban, A. T. Balaban, Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices, *J. Math. Chem.* **12** (1993) 309–318.
- 9 D. Plavšić, S. Nikolić, N. Trinajstić, Z. Mihalić, On the Harary index for the characterization of chemical graphs, *J. Math. Chem.* **12** (1993) 235–250.
- 10 Frank Harary (1921–2005), American mathematician.
- 11 B. Lučić, A. Miličević, S. Nikolić, N. Trinajstić, Harary index - twelve years later, *Croat. Chem. Acta* **75** (2002) 847–868.
- 12 K. Xu, M. Liu, K. C. Das, I. Gutman, B. Furtula, A survey on graphs extremal with respect to distance-based topological indices *MATCH Commun. Math. Comput. Chem.* **71** (2014) 461–508.
- 13 K. Xu, K. C. Das, N. Trinajstić, *The Harary Index of a Graph*, Springer, Heidelberg, 2015.
- 14 X. X. Li, Y. Z. Fan, The connectivity and the Harary index of a graph, *Discr. Appl. Math.* **181** (2015) 167–173.
- 15 B. Furtula, Atom-bond connectivity index versus Graovac-Ghorbani analog, *MATCH Commun. Math. Comput. Chem.* **75** (2016) 233–242.
- 16 A. T. Balaban, The Harary index of a graph, *MATCH Commun. Math. Comput. Chem.* **75** (2016) 243–245.
- 17 I. Gutman, B. Furtula, X. Li, Multicenter Wiener indices and their applications, *J. Serb. Chem. Soc.* **80** (2015) 1009–1017.
- 18 G. Chartrand, O. R. Oellermann, S. Tian, H. B. Zou, Steiner distance in graphs, *Časopis Pest. Mat.* **114** (1989) 399–410.
- 19 F. K. Hwang, D. S. Richards, P. Winter, *The Steiner Tree Problem*, North-Holland, Amsterdam, 1992.
- 20 E. Deza, M. M. Deza, *Dictionary of Distances*, Elsevier, Amsterdam, 2006.
- 21 M. M. Deza, E. Deza, *Encyclopedia of Distances*, Springer, Heidelberg, 2012.
- 22 Experimental physico-chemical data of octane isomers were taken from the site [www.moleculardescriptors.eu](http://www.moleculardescriptors.eu).
- 23 M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, 2000.

- 24 F. Shafiei, Relationship between thermodynamic properties and topological indices of the monocarboxylic acids applications in QSPR, *Iranian J. Math. Chem.* **6** (2015) 15–28.
- 25 R. Zanni, M. Gálvez-Llompарт, R. Garcia-Domenech, J. Gálvez, Latest advances in molecular topology applications for drug discovery, *Expert Opin. Drug Discov.* **10**(9) (2015) 1–13.