

Overall connectivity — a next generation molecular connectivity

Danail Bonchev

*Department of Marine Sciences and Program for Theory of Complex Systems, Texas A&M University,
 Ford Crockett Campus, 5007 Avenue U, Galveston, TX 77551, USA*

Received 31 October 2000; received in revised form 26 January 2001; accepted 26 January 2001

Dedicated to Milan Randić, Lemont Kier and Lowell Hall.

Abstract

The development of molecular connectivity concept and some of its key elements — Randić's inverse-square-root function and the detailed subgraph characterization — are analyzed. The concept of overall connectivity recently advanced is presented as a next step in unfolding the ideas of molecular connectivity by combining them with those of molecular complexity. Definitions of overall connectivity index, ϵ th-order overall connectivities, and overall connectivity vector are presented along with formulae for calculating these sets of topological indices for several classes of graphs of chemical relevance. Based on sums of adjacencies over all subgraphs (or up to a limiting subgraph size in large molecules), the overall connectivities increase both with molecule size and complexity, as expressed in branching and cyclicity of molecular skeleton. When applied to molecules containing heteroatoms, valence overall connectivities are constructed employing the Kier and Hall scheme. The usefulness of the novel indices is demonstrated by modeling physicochemical properties of alkane compounds. A detailed comparison is made with other models derived for the same set of compounds, proceeding from molecular connectivity, as well as with two other probe connectivity functions — the overall connectivity versions of the second Zagreb index, and a derivative inverse function of this index. The favorable comparisons indicate the need of molecular connectivity paradigm revisiting, and show the potential of the overall connectivity indices for QSPR/QSAR applications. © 2001 Elsevier Science Inc. All rights reserved.

Keywords: Overall connectivity; Molecular connectivity; Molecular complexity; QSPR/QSAR; Alkanes

1. Molecular connectivity and molecular complexity — the two fundamentals of overall connectivity

Twenty-five years ago several publications started a chain-reaction of interest toward molecular topology and its practical implications. Influenced by the Hückel molecular orbital theory, and by the classical book of Harary on graph theory [1], the atom–atom connectivity started being explored as connectivity of molecular graph. The connectivity of a graph G is described by its adjacency matrix $A(G)$, and the sum of all adjacency matrix entries a_{ij} is called [2] total adjacency $A(G)$:

$$A(G) = \sum_{ij} a_{ij} = \sum_i a_i \quad (1)$$

The second summation in Eq. (1) is over the vertex degrees (a_i), which count the vertex nearest neighbors and thus express the local (atomic) connectivity.

Total adjacency is a highly degenerate graph invariant. It has the same values for all molecules having the same number of bonds. To diminish strongly the degeneracy one

needs to make use of more distinctive function than the simple summation. Thus, the topological indices proposed by Gutman, Trinajstić et al. [3] in 1975 and called Zagreb indices, M_1 and M_2 :

$$M_1 = \sum_i a_i^2; \quad M_2 = \sum_{ij} a_i a_j \quad (2)$$

contain the sum of the squared vertex degrees (M_1), and the sum of vertex degrees products over all pairs of adjacent vertices (M_2), respectively.

Several months later, Randić [4] proposed his branching index $\chi(G)$, commonly noted nowadays as molecular connectivity index:

$$\chi(G) = \sum_{\text{all bonds}} (a_i a_j)^{-1/2} \quad (3)$$

The direct summation of vertex degree products of Eq. (2) has been changed in Eq. (3) to a summation of inverse-square-root terms. This specific function selection has been made to provide a high correlation of χ with properties of isomeric alkanes, thus showing the high sensitivity of the new molecular descriptor to variations in molecular structure.

E-mail address: bonchevd@aol.com (D. Bonchev).

At the end of the same year 1975, an essential development of the approach has been made by Kier et al. [5]. The summation of products of vertex degrees over all graph edges has been extended to analogical summation over products of vertex degrees of larger subgraphs, those with 2, 3, ..., k edges. Thus, a set of k th-order molecular connectivities, ${}^k\chi_c(G)$, has been defined:

$${}^k\chi_c(G) = \sum_{\text{all } k\text{-edge subgraphs}} (a_i a_j, \dots, a_{k+1})^{-1/2} \quad (4)$$

The subscript ‘c’ partitions further the set of subgraphs with a fixed number of edges into subgraph classes called path, cluster, path-cluster, etc. for which ‘c’ = p, c, and pc, respectively. This partitioning is an essential part of the established paradigm of molecular connectivity. It differs from the early attempts of Smolenski [6] in 1964, and Gordon [7] in 1972, which proposed to model molecular properties by a combination of individual contributions of each subgraph. The use of classes of subgraphs within molecular connectivity concept is a more practical approach to structure–property relationships. Interestingly, the intuitively made specific choice of substructure classes closely matches that developed later in polymer theory, which includes linear, star-like, comb-like, etc. molecular architectures [8].

The last important detail in the topological characterization of molecules by molecular connectivity concept was the introducing [9] of the zero-order term, ${}^0\chi$, which reduces the basic formula (2) to summation over vertices, not edges:

$${}^0\chi(G) = \sum_{\text{all vertices}} (a_i)^{-1/2} \quad (5)$$

This “size” term came to compensate for the two opposing trends of molecular connectivity indices: to increase with molecular size, and to decrease with molecular complexity (number, size, and central location of branches, cycles, etc.). Without this term, molecular connectivity indices would be well suited for isomeric molecules, but not very effective in series of molecules with a variable size. We will advocate in this paper that even this useful addition to the set of molecular connectivity indices does not compensate entirely for the opposing size and complexity trends, and a better solution will be searched by abandoning the inverse-square-root function ((3)–(5)). We shall not comment the further development of molecular connectivity concept, which introduced the valuable idea of valence connectivity [10], electrotopological state [11], and different attempts [12–15] to modify or extend the connectivity function, which are discussed in detail in other papers in this issue. Rather, we will focus on the search for an alternative measure of connectivity in molecules that would increase with both size and complexity. A brief introduction to molecular complexity would help to elucidate the approach.

In 1950s, attempts have been made to apply Shannon’s information theory [16] to evaluate molecular complexity [17,18] (see [2] for more details). Although the term

“complexity” was not explicitly used, the entire approach could be described as using the information content of a molecule as a measure of its complexity. Complexity of graphs was directly assessed on such basis by Mowshovitz [19] in 1968, and by Minoli [20] in 1976, the latter by proceeding from a combination of three simple graph invariants. Both measures were strongly degenerate, which prevented their usage for real chemical species.

The interest toward molecular complexity estimates has been gradually increasing after the pioneering work of Bertz [21,22] in 1981–1983. Bertz improved the topological information approach of Rashevsky [18] by considering more structure-sensitive graph substructures (two-edge subgraphs instead of vertices), as well as by adding a size term for dealing with series of compounds with a variable number of atoms. In 1987, Bonchev, Polansky [23] have shown that while the information theoretic measures describe well molecular elemental diversity, they are misleading in characterizing topological complexity of molecules. They introduced a concept according to which the higher molecular graph connectivity, the more complex the molecule. The connectivity measure used was the total adjacency $A(G)$ partitioned into contributions from edges included into four classes of subgraphs of increasing complexity: chains, bridges, branches, and cycles. Qualitatively a step in the right direction, this definition of complexity still needed a better quantitative formulation. Such a measure could well be the total count of subgraphs K in molecular graph, an idea mentioned briefly by Bertz and Herndon in 1985 in a paper devoted to molecular similarity analysis [24]. During 1996–1998, this idea was developed in detail simultaneously and independently by Bonchev [25–27] and Bertz and coworkers [28,29]. It is in [25–27] that the ideas of molecular connectivity and molecular complexity overlapped first to produce the new concept of overall connectivity.

2. From molecular connectivity to overall connectivity

Searching for a connectivity measure, which (unlike molecular connectivity indices) to increase with both molecular size and complexity, it was natural to proceed from the subgraph count K as a complexity measure, and to weight each subgraph with a simple connectivity graph-invariant. Following the “Ocam’s razor principle”, before looking for more complicated solutions, one may start with the simplest connectivity function, that of total adjacency, given in Eq. (1). Then, our overall connectivity concept can be simply formulated as

overall connectivity

= molecular complexity

+connectivity of all connected subgraphs

The overall connectivity thus appears as a meaningful measure of topological complexity of molecules (and, more

generally, of any structure independent of the physical nature of its elements), since it satisfies two fundamental requirements to a complexity measure: to increase with both the number of structural elements and their interconnectedness (see [30] for a more detailed introduction to molecular complexity, which is a more general concept than topological complexity).

More rigorously, the concept is defined as follows.

Definition 1. The overall connectivity $TC(G)$ of any graph G is defined as the sum of the total adjacencies $A_i(G_i)$ of all K subgraphs G_i of G having n_i vertices each:

$$TC(G) = \sum_{i=1}^K A_i(G_i \subset G) = \sum_{i=1}^K \sum_{j=1}^{n_i} a_{ij}(j \in G_i \subset G) \quad (6)$$

Definition 2. The e th-order overall connectivity ${}^eTC(G)$ of any graph G is defined as the sum of the total adjacencies $A_j({}^eG_j)$ of all eK subgraphs eG_j of G , which have e edges:

$${}^eTC(G) = \sum_{j=1}{{}^eK} A_i(G_j \subset G) \quad (7)$$

One can similarly define the e th-order overall connectivity ${}^eTC_k(G)$ of the k th class of subgraphs, eG_j of G , k standing for path, cluster, path-cluster, cycle, etc. by limiting the summing-up to the class of interest only.

Definition 3. The overall connectivity vector $TC'(G)$ of any graph G having q edges is the sequence of all eTC s listed in ascending order

$$TC'(G) = TC(^0TC, ^1TC, \dots, ^qTC) \quad (8)$$

or

$$TC'(G) = TC(^0TC, ^1TC, ^2TC, ^3TC_k, \dots, ^qTC_k) \quad (9)$$

Definition 4. The overall connectivity $TC(G)$ is defined in two versions, depending on whether vertex degrees are those of the entire graph G or those of the subgraph G_i :

$$TC(G) = \sum_{i=1}^K A_i(a_i \in G);$$

$$TC1(G) = \sum_{i=1}^K A_i(a_i \in G_i) \quad (10)$$

Definition 5. Complexity index K can be presented in a similar manner as a vector $K'(G)$:

$$K'(G) = K(^0K, ^1K, ^2K, \dots, ^qK) \quad (11)$$

$$K(G) = \sum_{e=0}^q {}^eK \quad (12)$$

where eK is the number of subgraphs having e edges.

Definition 6. Cumulative p th-order overall connectivity, ${}^pTC'(G)$, and complexity index, ${}^pK'(G)$, are defined as:

$${}^pTC'(G) = \sum_{e=0}^p {}^eTC(G) \quad (13)$$

$${}^pK'(G) = \sum_{e=0}^p {}^eK(G) \quad (14)$$

These indices could be used as additional descriptors or in case of large molecules, for which the calculation of K , TC and $TC1$ would require large amount of computer time.

Definition 7. The valence overall connectivity $TC^v(G)$ of any graph G is defined as the sum of the valence total adjacencies $A_i^v(G_i)$ of all K subgraphs G_i of G having N_i vertices each.

$$TC^v(G) = \sum_{i=1}^K a_i^v(G_i \subset G) = \sum_{i=1}^K \sum_{j=1}^{N_i} A_{ij}^v(j \in G_i \subset G) \quad (15)$$

where $a_j^v = \delta^v$ is the valence term of Kier and Hall [10], which replaces the vertex degree a_j .

The novel elements of overall connectivity concept, as defined above, are the following.

1. The use of cumulative indices (TC , $TC1$, K), which increase with size and topological complexity, as presented by branching and cyclicity patterns of molecular skeleton. Such an index, if introduced within molecular connectivity scheme, would be of little use, because of the opposing trends of increasing with size and decreasing with complexity. The ordering of isomeric acyclic and cyclic structures of increasing complexity produced by the overall connectivity indices is analyzed elsewhere [31]. A detailed comparison of the ordering produced by different complexity measures is given in [32,33].
2. More general e th-order classes are introduced along with the more detailed classes of paths, clusters, path-clusters, etc. used within molecular connectivity concept.
3. A vector-type distribution is constructed with the e th-order values presenting as partitioning of the cumulative index. The overall connectivity vectors (Eqs. (8), (9) and (11)) can be used for additional discrimination in case of degeneracy of the cumulative index.
4. A second type of connectivity indices ($TC1$) is introduced, for which the vertex degrees are taken as they are in the subgraphs but not in the entire graph. This makes the set of e th-order $TC1$ indices independent of the specific molecule they are incorporated in, and a standard table with the eTC1 and eK values of the low-order subgraphs can be used for any kind of compounds. Such tables for acyclic and cyclic subgraphs are given in [27].

Table 1

Path, cluster, and path–cluster values of the overall connectivity indices TC and TC1 in C3–C8 alkanes^a

No.	Molecule	Indices	⁰ X	¹ X	² X	³ X _p	³ X _c	⁴ X _p	⁴ X _c	⁴ X _{pc}	⁵ X _p	⁵ X _c	⁵ X _{pc}	⁶ X _p	⁶ X _c	⁶ X _{pc}	X
1	C3	TC1	0	4	4												8
		TC	4	6	4												14
2	<i>n</i> -C4	TC1	0	6	8	6	0										20
		TC	6	10	10	6	0										32
3	2MC3	TC1	0	6	12	0	6										24
		TC	6	12	15	0	6										39
4	<i>n</i> -C5	TC1	0	8	12	12	0	8	0	0							40
		TC	8	14	16	14	0	8	0	0							60
5	MC4	TC1	0	8	16	12	6	0	0	8							50
		TC	8	16	23	14	7	0	0	8							76
6	22MC3	TC1	0	8	24	0	24	0	8	0							64
		TC	8	20	36	0	28	0	8	0							100
7	<i>n</i> -C6	TC1	0	10	16	18	0	16	0	0	10	0	0				70
		TC	10	18	22	22	0	18	0	0	10	0	0				100
8	2MC5	TC1	0	10	20	18	6	16	0	8	0	0	10				88
		TC	10	20	29	24	7	18	0	9	0	0	10				127
9	3MC5	TC1	0	10	20	24	6	8	0	16	0	0	10				94
		TC	10	20	31	30	8	9	0	18	0	0	10				136
10	23MC4	TC1	0	10	24	24	12	0	0	32	0	10	0				112
		TC	10	22	38	32	16	0	0	36	0	10	0				164
11	22MC4	TC1	0	10	28	18	24	0	8	24	0	0	10				122
		TC	10	24	46	24	31	0	9	27	0	0	10				181
12	<i>n</i> -C7	TC1	0	12	20	24	0	24	0	0	20	0	0	12	0	0	112
		TC	12	22	28	30	0	28	0	0	22	0	0	12	0	0	154
13	2MC6	TC1	0	12	24	24	6	24	0	8	20	0	10	0	0	12	140
		TC	12	24	35	32	7	30	0	9	22	0	11	0	0	12	194
14	3MC6	TC1	0	12	24	30	6	24	0	16	10	0	20	0	0	12	154
		TC	12	24	37	40	8	29	0	19	11	0	22	0	0	12	214
15	3EC5	TC1	0	12	24	36	6	24	0	24	0	0	30	0	0	12	168
		TC	12	24	39	48	9	27	0	30	0	0	33	0	0	12	234
16	24MMC5	TC1	0	12	28	24	12	32	0	16	0	0	40	0	0	12	176
		TC	12	26	42	36	14	40	0	20	0	0	44	0	0	12	246
17	23MMC5	TC1	0	12	28	36	12	16	0	40	0	10	30	0	0	12	196
		TC	12	26	46	50	17	20	0	49	0	11	33	0	0	12	276
18	22MC5	TC1	0	12	32	24	24	24	8	24	0	0	40	0	0	12	200
		TC	12	28	52	36	31	30	9	30	0	0	44	0	0	12	284
19	33MC5	TC1	0	12	32	36	24	8	8	48	0	0	40	0	0	12	220
		TC	12	28	56	50	34	10	10	58	0	0	44	0	0	12	314
20	223MC5	TC1	0	12	36	36	30	0	8	72	0	10	40	0	12	0	256
		TC	12	30	63	54	43	0	10	90	0	11	44	0	12	0	369
21	<i>n</i> -C8	TC1	0	14	24	30	0	32	0	0	30	0	0	24	0	0	168
		TC	14	26	34	38	0	38	0	0	34	0	0	26	0	0	224
22	2MC7	TC1	0	14	28	30	6	32	0	8	30	0	10	24	0	12	208
		TC	14	28	41	40	7	40	0	9	36	0	11	26	0	13	279
23	3MC7	TC1	0	14	28	36	6	32	0	16	30	0	20	12	0	24	232
		TC	14	28	43	48	8	41	0	19	35	0	23	13	0	26	312
24	4MC7	TC1	0	14	28	36	6	40	0	16	20	0	30	12	0	24	240
		TC	14	28	43	50	8	49	0	20	24	0	34	13	0	26	323
25	3EC6	TC1	0	14	28	42	6	40	0	24	20	0	40	0	0	36	264
		TC	14	28	45	58	9	49	0	31	22	0	47	0	0	39	356

Table 1 (Continued)

No.	Molecule	Indices	⁰ X	¹ X	² X	³ X _p	³ X _c	⁴ X _p	⁴ X _c	⁴ X _{pc}	⁵ X _p	⁵ X _c	⁵ X _{pc}	⁶ X _p	⁶ X _c	⁶ X _{pc}	X
26	25MC6	TC1	0	14	32	30	12	32	0	16	40	0	20	0	0	48	258
		TC	14	30	48	42	14	44	0	18	48	0	24	0	0	52	348
27	24MC6	TC1	0	14	32	36	12	40	0	24	20	0	50	0	0	48	290
		TC	14	30	50	52	15	53	0	30	24	0	59	0	0	52	393
28	23MC6	TC1	0	14	32	42	12	32	0	40	20	10	40	0	0	48	304
		TC	14	30	52	60	17	42	0	50	24	11	48	0	0	52	414
29	22MC6	TC1	0	14	36	30	24	32	8	24	30	0	40	0	0	48	300
		TC	14	32	58	44	31	44	9	30	36	0	47	0	0	52	411
30	34MC6	TC1	0	14	32	48	12	32	0	48	10	10	60	0	0	48	328
		TC	14	30	54	68	18	42	0	62	12	12	70	0	0	52	448
31	33MC6	TC1	0	14	36	42	24	32	8	48	10	0	70	0	0	48	346
		TC	14	32	62	62	34	42	10	61	12	0	82	0	0	52	477
32	23MEC5	TC1	0	14	32	48	12	40	0	48	0	10	70	0	0	48	336
		TC	14	30	54	70	18	49	0	64	0	12	82	0	0	52	459
33	33MEC5	TC1	0	14	36	54	24	24	8	72	0	0	90	0	0	48	384
		TC	14	32	66	78	37	30	11	93	0	0	105	0	0	52	532
34	234MC5	TC1	0	14	36	48	18	32	0	64	0	20	80	0	0	60	386
		TC	14	32	61	72	26	44	0	84	0	24	96	0	0	65	532
35	224MC5	TC1	0	14	40	30	30	48	8	32	0	0	100	0	0	60	376
		TC	14	34	65	50	38	66	9	44	0	0	120	0	0	65	519
36	223MC5	TC1	0	14	40	48	30	24	8	80	0	30	80	0	12	48	428
		TC	14	34	71	74	44	33	10	107	0	36	95	0	13	52	597
37	233MC5	TC1	0	14	40	54	30	16	8	96	0	30	80	0	12	48	442
		TC	14	34	73	82	46	22	11	126	0	35	96	0	13	52	618
38	2233MC4	TC1	0	14	48	54	48	0	16	144	0	90	60	0	72	0	560
		TC	14	38	90	90	74	0	22	198	0	108	72	0	78	0	798

^a For brevity, TC and TC1 are denoted as X.

In this paper we present (vide infra) in Table 1 such values for the ^eTC1_k and ^eK_k indices of the path, cluster, and path–cluster subsets of subgraphs. The second-type connectivity indices could be introduced for any specific connectivity function, including the inverse-square-root function of Randić (Eq. (3)).

- The simple additive connectivity function used enables the deriving of analytical formulae for some standard classes of structures.

3. Formulae for overall connectivity and complexity indices of some classes of graphs

Paths (P_n):

$${}^eK(P_n) = n - e; \quad K(P_n) = \frac{n(n+1)}{2} \quad (16)$$

$${}^e\text{TC1}(P_n) = 2e(n - e);$$

$$\text{TC1}(P_n) = \frac{n(n-1)(n+1)}{3} = 2W(P_n) \quad (17)$$

$${}^e\text{TC}(P_n) = 2[q(e+1) - e^2];$$

$$\text{TC}(P_n) = \frac{n(n-1)(n+4)}{3} \quad (18)$$

Cycles (C_n):

$${}^eK(C_n) = n - e; \quad {}^qK(C_n) = 1;$$

$$K(C_n) = n^2 + 1 \quad (19)$$

$${}^e\text{TC1}(C_n) = 2ne; \quad {}^q\text{TC1}(C_n) = 2n;$$

$$\text{TC1}(C_n) = n(n^2 - n + 2) \quad (20)$$

$${}^e\text{TC}(C_n) = 2n(e+1); \quad {}^q\text{TC}(C_n) = 2n;$$

$$\text{TC}(C_n) = n(n^2 + n + 2) \quad (21)$$

Star graphs (S_n):

$${}^eK(S_n) = \binom{q}{e}; \quad {}^0K(S_n) = n;$$

$$K(S_n) = 2^q + q \quad (22)$$

$${}^e\text{TC1}(S_n) = 2e \binom{q}{e}; \quad \text{TC1}(S_n) = 2^q q \quad (23)$$

$${}^e\text{TC}(S_n) = (q + e) \binom{q}{e}; \quad {}^0\text{TC}(S_n) = 2q;$$

$$\text{TC}(S_n) = 2^q + \sum_{e=1}^q (q + e) \binom{q}{e} \quad (24)$$

As shown in Eq. (17), the TC1 index of linear graphs is related to their Wiener number by a factor of two.

4. Overall connectivity indices of C3–C8 alkanes

Table 1 presents the values of the connectivity indices TC, and TC1, as well as those of ${}^e\text{TC1}_k$ and ${}^e\text{TC}_k$ of 38 alkane molecules having three to eight carbon atoms. The e th-order values are those of the respective path, cluster, and path–cluster subgraphs. The values of the cumulative eK , ${}^e\text{TC1}$ and ${}^e\text{TC}$ indices of these structures were given in our previous publications [27,34] along with such data for some cyclic structures. The structure names in Table 1 are abbreviated with the number of carbon atoms in the longest chain, with a prefix showing methyl (M) or/and ethyl (E) branches, and with numbers showing branch position.

Quantitative structure–property relationships with overall connectivity indices. Comparison of four connectivity functions.

The applicability of the overall connectivity indices to QSPR studies was verified by linear regression modeling of 10 physicochemical properties of C3–C8 alkanes, used as a benchmark: boiling points [35] (with two octane and four nonane boiling points corrected by Herndon [36], who verified all alkane boiling points against those of the original measurements), T_B in $^{\circ}\text{C}$; critical temperatures [35], T_c in $^{\circ}\text{C}$; critical pressures [35], P_c in atm; critical volume [35], V_c in l/mol [37]; molar volume, V_m in cm^3/mol [9]; molecular refraction, R_m in cm^3/mol [9]; surface tension [35], ST in dyn/cm ; the heat of formation in gaseous state, $\Delta H_f(g)$ in kJ/mol [38]; the heat of vaporization, ΔH_v in kJ/mol [38]; the heat of atomization, ΔH_a in kcal/mol [9]. The values of the molecular connectivity indices are taken from Kier and Hall's monograph [9]. The calculation was performed by using from the OASIS software [39] the option that selects the best models with one to five parameters after generating all models with the given sets of variables (a method equivalent to all possible subsets regression), and eliminating the insignificant ones. The selection for TC, TC1, and χ was from about 3000 models, whereas for the combined sets of TC, TC1, and K , more than 250,000 models were compared. The model comparison was done by using correlation coefficient, r , standard deviation, s , Fisher's ratio, F , and the average cross-validated standard deviation, $s(\text{cv})$. The latter was also used as a criterion for elimination of chance correlations.

Our previous calculations of the same sets of compounds and properties with e th-order overall connectivities have already shown that these topological descriptors compare

favorably with molecular connectivity indices [27,34]. The models obtained in the present study with the more detailed path, cluster and path–cluster terms, ${}^e\text{TC}_k$, ${}^e\text{TC1}_k$, and eK_k , showed a further improvement for some properties. Thus, the standard deviations of our previous five-variable models of heat of formation, critical temperature, and critical pressure scaled down from 1.23 to 1.02 kJ/mol , from 4.7 to 3.25 $^{\circ}\text{C}$, and from 0.56 to 0.37 atm, respectively. The second-order overall connectivity indices appeared most frequently in the models, followed by the first-order ones. As shown in Table 2, all five-variable models with overall connectivities show lower standard deviations than those produced by molecular connectivity indices, the difference being considerable for heats of atomization, boiling point, and critical temperature.

The comparison with molecular connectivity indices questioned the key element of molecular connectivity paradigm — the usage of the inverse-square-root function (Eqs. (3) and (4)). It was of interest to check whether this function was an improvement over its predecessor, the second Zagreb index M_2 (Eq. (2)), which like overall connectivity indices increases with both size and complexity. The Zagreb indices M_1 and M_2 indices were extended very recently to overall-type indices as brief remarks simultaneously and independently by Nikolić et al. [32] and by Galbraikh et al. [40] and in full detail by Bonchev and Trinajstić [41]. The overall OM2 index is defined below as a sum over all e th-order OM2 indices with $e = 0, 1, 2, \dots, q$. These terms can be defined either generally (Eq. (26a)) or separately for the classes of paths, clusters, path–clusters, etc. (Eq. (26b)):

$$\text{OM2}(G) = \sum_{e=0}^q {}^e\text{OM2}(G_i \subset G) \quad (25)$$

$${}^e\text{OM2}(G) = \sum_{\text{all } e\text{-edge subgraphs}} (a_i a_j, \dots, a_{e+1}) \quad (26a)$$

$${}^e\text{OM2}_k(G) = \sum_{\text{all } e\text{-edge subgraphs of type } k} (a_i a_j, \dots, a_{e+1}) \quad (26b)$$

A zero-order OM2-index was introduced by analogy with molecular connectivity zero-order term, i.e. by applying the connectivity function to vertices instead of edges. This made the zero-order OM2-index identical with zero-order overall connectivity TC index, what partially explains the good performance of the set of ${}^e\text{OM2}$ indices shown below.

Besides comparing the extended second Zagreb index OM2 to molecular connectivity χ , it was also of interest to check what would be the performance of a connectivity function intermediate between these two. Such a function, which we denote by ON, has terms inverse to those of M_2 but unlike molecular connectivity indices they do not include a square-root:

Table 2

The best overall connectivity models vs. the best molecular connectivity models for 10 alkane properties^a

Property	Parameters	<i>r</i>	<i>s</i>	<i>F</i>	<i>s</i> (cv)
Five-variable models					
<i>T_B</i>	⁰ <i>K</i> , ¹ <i>TC</i> , ³ <i>TC_c</i> , ⁴ <i>TC1_p</i> , ⁵ <i>TC_p</i>	0.9993	1.60	4914	2.00
	⁰ <i>χ</i> , ² <i>χ</i> , ³ <i>χ_c</i> , ⁴ <i>χ_c</i> , ⁵ <i>χ_{pc}</i>	0.9972	3.31	1151	4.06
ΔH_f (g)	² <i>K</i> , ³ <i>TC_c</i> , ⁴ <i>TC_c</i> , ⁵ <i>TC_{pc}</i> , ⁵ <i>TC1_{pc}</i>	0.9995	1.02	6324	1.26
	⁰ <i>χ</i> , ¹ <i>χ</i> , ² <i>χ</i> , ⁵ <i>χ_c</i> , ⁵ <i>χ_{pc}</i>	0.9991	1.37	3479	1.69
ΔH_v	<i>TC1</i> , ¹ <i>TC</i> , ² <i>TC</i> , ³ <i>TC_p</i> , ⁴ <i>TC1_c</i>	0.9950	0.67	633	0.81
	⁰ <i>χ</i> , ² <i>χ</i> , ³ <i>χ_p</i> , ³ <i>χ_c</i> , ⁴ <i>χ_c</i>	0.9930	0.79	453	0.97
ΔH_a	¹ <i>TC</i> , ² <i>TC1</i> , ³ <i>K</i> , ⁵ <i>TC_{pc}</i> , ⁵ <i>TC1_{pc}</i>	1.0000	0.30	11900000	0.39
	⁰ <i>χ</i> , ² <i>χ</i> , ³ <i>χ_p</i> , ⁴ <i>χ_p</i> , ⁴ <i>χ_c</i>	0.9999	5.78	32400	7.45
<i>ST</i>	<i>K</i> , <i>TC1</i> , ² <i>K</i> , ⁴ <i>K</i> , ⁵ <i>TC1_c</i>	0.9963	0.17	727	0.22
	¹ <i>χ</i> , ² <i>χ</i> , ³ <i>χ_c</i> , ⁴ <i>χ_{pc}</i> , ⁵ <i>χ_{pc}</i>	0.9938	0.22	432	0.27
<i>R_m</i>	⁰ <i>K</i> , ² <i>K</i> , ³ <i>TC_p</i> , ⁵ <i>K</i> , ⁵ <i>TC_{pc}</i>	1.0000	0.041	70200	0.47
	¹ <i>χ</i> , ² <i>χ</i> , ³ <i>χ_p</i> , ⁴ <i>χ_{pc}</i> , ⁴ <i>χ_c</i>	1.0000	0.044	60700	0.053
<i>V_m</i>	¹ <i>TC1</i> , ² <i>TC</i> , ³ <i>TC_p</i> , ³ <i>TC1_c</i> , ⁴ <i>TC_c</i>	0.9998	0.33	11400	0.42
	¹ <i>χ</i> , ² <i>χ</i> , ⁴ <i>χ_p</i> , ⁴ <i>χ_{pc}</i> , ⁴ <i>χ_c</i>	0.9997	0.36	9955	0.48
<i>P_c</i>	¹ <i>TC1</i> , ³ <i>TC_p</i> , ⁴ <i>TC_c</i> , ⁴ <i>TC_{pc}</i> , ⁶ <i>TC_c</i>	0.9955	0.37	705	0.55
	⁰ <i>χ</i> , ¹ <i>χ</i> , ⁴ <i>χ_{pc}</i> , ⁵ <i>χ_{pc}</i> , ⁶ <i>χ_{pc}</i>	0.9920	0.50	397	0.69
<i>V_c</i>	⁰ <i>K</i> , ¹ <i>TC1</i> , ³ <i>TC_p</i> , ⁴ <i>TC_c</i> , ⁴ <i>TC1_c</i>	0.9948	0.0079	612	0.0125
	⁰ <i>χ</i> , ² <i>χ</i> , ³ <i>χ_p</i> , ⁴ <i>χ_p</i> , ⁶ <i>χ_p</i>	0.9936	0.0087	497	0.0103
<i>T_c</i>	<i>TC1</i> , ⁰ <i>K</i> , ² <i>TC1</i> , ⁴ <i>TC1_p</i> , ⁵ <i>TC_{pc}</i>	0.9982	3.25	1790	3.99
	⁰ <i>χ</i> , ¹ <i>χ</i> , ³ <i>χ_p</i> , ³ <i>χ_c</i> , ⁶ <i>χ_c</i>	0.9962	4.76	831	8.52

^a Statistics comparison of five-parameter models. The cross-validation correlation coefficient $q = r$ for all TC and χ five-parameter models.

$$\text{ON}(G) = \sum_{e=0}^q {}^e\text{ON}(G_i \subset G) \quad (27)$$

$${}^e\text{ON}(G) = \sum_{\text{all } e\text{-edge subgraphs}} (a_i a_j, \dots, a_{e+1})^{-1} \quad (28a)$$

$${}^e\text{ON}_k(G) = \sum_{\text{all } e\text{-edge subgraphs of type } k} (a_i a_j, \dots, a_{e+1})^{-1} \quad (28b)$$

Thus, a systematic comparison of four connectivity functions was enabled starting with the simple summation over vertex degrees, and continuing with the product, inverse product, and inverse-square-root product of vertex degrees in all subgraphs:

$$\begin{aligned} \text{TC, TC1} \left(\sum_i a_i \right) &\rightarrow \text{OM2} \left(\sum_{ij} a_i a_j \right) \\ &\rightarrow \text{ON} \left(\sum_{ij} (a_i a_j)^{-1} \right) \\ &\rightarrow \chi \left(\sum_{ij} (a_i a_j)^{-1/2} \right) \end{aligned} \quad (29)$$

Before presenting the results of this comparison, we give in Table 3 the values of the ${}^e\text{ON}_k$ and ON indices for the se-

lected series of C3–C8 alkane compounds. The corresponding values of OM2 indices are given elsewhere [41].

The methodology of our comparison was to deal with real cases of a mixture of nonisomeric and isomeric compounds. Dealing with isomers only is usually recommended to avoid the high correlation due to the number of atoms typical for homologous series of compounds. Another argument in favor of test sets of isomeric compounds is that such a choice enables the identification of descriptors, which are highly sensitive to variations in molecular topology. However, models derived from an isomeric series of compounds have limited practical value, even when one tries to apply models derived from one isomeric series to the closest other isomeric series. Thus, models of physicochemical properties derived for octanes do not hold for nonanes and decanes [42]. On the other hand, the statistical parameter of practical value in developing new materials and drugs is standard deviation; correlation coefficient plays only a secondary role.

In support of our concept for testing topological indices we have compared the performance of the C3–C8 models and the C8 models in predicting the boiling points and molar volumes of C9 alkanes.

The data sources we used provided experimental data for all 35 nonane boiling points [35,36] and 14 nonane molar volumes (the only ones given in [9]). The comparison for each property was made as follows. First, the best overall connectivity-type model (with the more general ${}^e\text{TC}$, ${}^e\text{TC1}$, and eK terms but not the detailed path, cluster and

Table 3

C3–C8 alkane values of the overall index ON^a

No.	Molecule	⁰ N	¹ N	² N	³ N _p	³ N _c	⁴ N _p	⁴ N _c	⁴ N _{pc}	⁵ N _p	⁵ N _c	⁵ N _{pc}	⁶ N _p	⁶ N _c	⁶ N _{pc}	ON
1	C3	2.5	1	0.5												4
2	n-C4	3	1.25	0.5	0.25	0										5
3	2MC3	3.333	1	1	0	0.333										5.667
4	n-C5	3.5	1.5	0.625	0.25	0	0.125	0	0							6
5	MC4	3.833	1.333	0.833	0.333	0.167	0	0	0.167							6.667
6	22MC3	4.25	1	1.5	0	1	0	0.25	0							8
7	n-C6	4	1.75	0.75	0.313	0	0.125	0	0	0.063	0	0				7
8	2MC5	4.333	1.583	1	0.25	0.167	0.167	0	0.083	0	0	0.083				7.667
9	3MC5	4.333	1.667	0.75	0.5	0.083	0.083	0	0.167	0	0	0.083				7.667
10	23MC4	4.667	1.444	1.111	0.444	0.222	0	0	0.444	0	0.111	0				8.444
11	22MC4	4.75	1.375	1.25	0.375	0.625	0	0.125	0.375	0	0	0.125				9
12	n-C7	4.5	2	0.875	0.375	0	0.156	0	0	0.063	0	0	0.031	0	0	8
13	2MC6	4.833	1.833	1.125	0.333	0.167	0.125	0	0.083	0.083	0	0.042	0	0	0.042	8.667
14	3MC6	4.833	1.917	0.917	0.458	0.083	0.167	0	0.125	0.042	0	0.083	0	0	0.042	8.667
15	3EC5	4.833	2	0.75	0.5	0.042	0.25	0	0.125	0	0	0.125	0	0	0.042	8.667
16	24MMC5	5.167	1.667	1.389	0.222	0.333	0.222	0	0.111	0	0	0.222	0	0	0.056	9.389
17	23MMC5	5.167	1.778	1.056	0.556	0.167	0.111	0	0.333	0	0.056	0.167	0	0	0.056	9.444
18	22MC5	5.25	1.625	1.438	0.25	0.625	0.188	0.125	0.188	0	0	0.25	0	0	0.063	10
19	33MC5	5.25	1.75	1.063	0.625	0.375	0.063	0.063	0.5	0	0	0.25	0	0	0.063	10
20	223MC5	5.583	1.5	1.5	0.5	0.583	0	0.083	0.75	0	0.083	0.333	0	0.083	0	11
21	n-C8	5	2.25	1	0.438	0	0.188	0	0	0.078	0	0	0	0.031	0	9
22	2MC7	5.333	2.083	1.25	0.396	0.167	0.167	0	0.083	0.063	0	0.042	0.042	0	0.021	9.667
23	3MC7	5.333	2.167	1.042	0.542	0.083	0.146	0	0.125	0.083	0	0.063	0.021	0	0.042	9.667
24	4MC7	5.333	2.167	1.083	0.417	0.083	0.271	0	0.083	0.042	0	0.104	0.021	0	0.042	9.667
25	3EC6	5.333	2.25	0.917	0.5	0.042	0.25	0	0.104	0.083	0	0.104	0	0	0.063	9.667
26	25MC6	5.667	1.917	1.5	0.361	0.333	0.111	0	0.167	0.111	0	0.056	0	0	0.111	10.362
27	24MC6	5.667	2	1.306	0.444	0.25	0.194	0	0.167	0.056	0	0.167	0	0	0.111	10.390
28	23MC6	5.667	2.028	1.222	0.528	0.167	0.167	0	0.306	0.056	0.056	0.111	0	0	0.111	10.444
29	22MC6	5.75	1.875	1.563	0.344	0.625	0.125	0.125	0.188	0.094	0	0.156	0	0	0.125	11
30	34MC6	5.667	2.111	1	0.694	0.111	0.167	0	0.278	0.028	0.028	0.222	0	0	0.111	10.444
31	33MC6	5.75	2	1.25	0.531	0.375	0.188	0.063	0.375	0.031	0	0.281	0	0	0.125	11
32	2M3EC5	5.667	2.111	1.083	0.5	0.139	0.306	0	0.222	0	0.028	0.25	0	0	0.111	10.444
33	3M3EC5	5.75	2.125	0.938	0.75	0.219	0.188	0.031	0.469	0	0	0.375	0	0	0.125	11
34	234MC5	6	1.889	1.370	0.593	0.259	0.148	0	0.444	0	0.074	0.296	0	0	0.185	11.295
35	224MC5	6.083	1.708	1.833	0.208	0.792	0.25	0.125	0.167	0	0	0.417	0	0	0.208	11.833
36	223MC5	6.083	1.833	1.458	0.542	0.583	0.125	0.083	0.542	0	0.125	0.375	0	0.042	0.167	12
37	233MC5	6.083	1.875	1.333	0.708	0.375	0.083	0.042	0.75	0	0.167	0.333	0	0.042	0.167	12
38	2233MC4	6.5	1.563	1.875	0.563	0.875	0	0.125	1.125	0	0.563	0.375	0	0	0.375	14

^a For brevity, ON is denoted as *N*.

Table 4

Comparison of the C3–C8 and C8 models predictivity for boiling points and molar volumes of nonanes

Training set	Variables	<i>n</i>	<i>r</i>	<i>F</i>	<i>s</i>
Boiling points					
C3–C8	⁰ K, TC, TC1, ² TC1, ³ K	38	0.9992	4161	1.74
C3–C9 (prediction)	⁰ K, TC, TC1, ² TC1, ³ K	73	0.9988	5436	2.08
C3–C9 (best model)	TC, TC1, ¹ TC1, ² TC1, ⁵ TC	73	0.9988	5707	2.03
C8	K, TC, ² TC, ³ TC1	18	0.9866	119	1.14
C8–C9 (prediction)	K, TC, ² TC, ³ TC1	53	0.8920	47	6.00
C8–C9 (best model)	TC, TC1, ¹ TC, ² K	53	0.9874	468	2.10
Molar volumes ^a					
C3–C8	⁰ K, ¹ TC, ² TC, ⁴ K, ⁷ K	32	0.9998	13600	0.31
C3–C9 (prediction)	⁰ K, ¹ TC, ² TC, ⁴ K, ⁷ K	46	0.9994	6927	0.62
C3–C9 (best model)	⁰ K, TC, TC1, ² TC, ³ TC1	46	0.9996	9960	0.52
C8	³ K, ³ TC1, ⁵ TC, ⁵ TC1	16	0.9971	480	0.23
C8–C9 (prediction)	³ K, ³ TC1, ⁵ TC, ⁵ TC1	30	0.7653	8.8	5.70
C8–C9 (best model)	⁰ K, ² TC1, ⁴ K, ⁵ TC	30	0.9981	1646	0.54

^a The 16 octane molar volumes and 14 nonane ones are given in [9].

Table 5
Standard deviations of the best five-parameter models with different connectivity functions

Property	Standard deviation, <i>s</i>
Boiling point	OC ^a (1.60) < TC1 (1.70) < ON (2.53) < OM2 (2.71) < TC (2.81) < χ (3.31)
Heat of formation	OC (1.02) < TC (1.19) < TC1 (1.33) < χ (1.37) < OM2 (1.55) < ON (1.87)
Heat of vaporization	OC (0.67) < TC (0.70) = OM2 (0.70) < TC1 (0.75) < ON (0.78) < χ (0.79)
Heat of atomization	OC (0.30) < TC1 (0.34) < OM2 (0.39) < ON (2.8) < TC (5.3) < χ (5.8)
Surface tension	OC (0.17) < TC (0.21) = TC1 (0.21) < χ (0.22) < OM2 (0.27) = ON (0.27)
Molecular refraction	OC (0.041) = TC (0.041) < χ (0.044) < TC1 (0.047) < OM2 (0.050) < ON (0.066)
Molar volume	OC (0.33) < TC (0.35) < χ (0.36) < ON (0.41) < TC1 (0.48) < OM2 (0.57)
Critical volume	OM2 (0.0076) < OC (0.0079) < ON (0.0080) < TC (0.0081) < χ (0.0087) < TC1 (0.0089)
Critical pressure	OC (0.37) = TC1 (0.37) < TC (0.38) < OM2 (0.43) < χ (0.50) < ON (0.66)
Critical temperature	OC (3.25) < TC1 (4.40) < ON (4.57) < TC (4.64) < χ (4.76) < OM2 (5.13)

^a OC (abbreviation for overall connectivity) is used for the set of TC + TC1 + *K*.

Table 6
Summary of the comparative performance of the five-parameter linear models produced by overall connectivity and molecular connectivity functions

Index	Rank				
	1	2	3	4	5
TC	3 + 2 ^a	1	2	2	
TC1	4 + 1 ^a	1	2	1	1
OM2	1 + 1 ^a	1	2	2 + 1 ^b	2
ON		3	2	1 + 1 ^b	3
χ		2	2	3	3
OC (TC + TC1 + <i>K</i>)	9	1			

^a Shared rank 1.

^b Shared rank 4.

path-cluster terms, ^eTC_{*k*}, ^eTC1_{*k*}, and ^eK_{*k*}, used in the rest of this study) was derived for the C3–C8 series and, separately, for the C8 series; due to the different size of these data sets the models included five and respectively four variables. Then, with the same variables, a new regression was obtained for the extended series, which includes nonanes, and the statistics of the models were compared. As shown in Table 4, the extended models based on the C3–C9 series of compounds are characterized by almost identical correlation coefficients with those of the initial C3–C8 models. They predict the values of the two nonane properties with a

reasonably small increase in the standard deviation (0.34°C for boiling point and 0.31 l/mol for molar volume). These statistical parameters are close to those of the model independently obtained by testing all variable combinations in the extended series. In contrast, applying the C8 models to the extended C8 + C9 series leads to disastrous results: the correlation coefficient drops by about 0.10 for boiling points and by more than 0.20 for molar volumes; the standard deviation skyrockets for boiling points from 1.14 to 6.00°C, and for molar volume from 0.23 to 5.70 l/mol.

Coming back to our goal to compare the performance of the four connectivity functions given in Eq. (29), we now proceed with the C3–C8 set of compounds and compare the best models obtained with one to five variables for each of the 10 physicochemical properties. Table 5 shows only the most relevant five-variable models, which provide the smallest standard deviation. The two versions of overall connectivity index, ^eTC and ^eTC1, are examined both separately, and together in a common set of descriptors with complexity indices, ^eK. The performance of the connectivity functions tested is summarized in Table 6. The specific descriptors of TC-, TC1-, ON-, and OM2-type, which take part in the models of Table 5, are listed in Table 7 (those for the combined overall connectivity, and for molecular connectivity models were given in Table 2).

Table 7
Parameters included in the best five-parameter models with connectivity functions *X* = TC, TC1, OM2, and ON^a

Property	TC	TC1	OM2	ON
<i>T</i> _B	¹ X, ³ X _p , ³ X _c , ⁴ X _p , ⁴ X _c	² X, ³ X _c , ⁴ X _p , ⁴ X _c , ⁵ X _p	⁰ X, ¹ X, ² X, ³ X _c , ⁴ X _p	⁰ X, ¹ X, ² X, ³ X _p , ⁴ X _p
ΔH_f (g)	¹ X, ² X, ⁵ X _c , ⁵ X _{pc} , ⁶ X _p	² X, ³ X _c , ⁴ X _c , ⁴ X _{pc} , ⁵ X _{pc}	¹ X, ² X, ³ X _p , ³ X _c , ⁵ X _c	⁰ X, ¹ X, ² X, ⁵ X _c , ⁶ X _{pc}
ΔH_v	¹ X, ¹ X, ² X, ³ X _p , ⁴ X _c	² X, ³ X _c , ⁴ X _c , ⁵ X _p , ⁶ X _{pc}	⁰ X, ¹ X, ² X, ⁴ X _{pc}	⁰ X, ¹ X, ² X, ³ X _p , ⁴ X _p
ΔH_a	¹ X, ³ X _p , ³ X _c , ⁴ X _c , ⁵ X _{pc}	¹ X, ² X, ³ X _p , ³ X _c , ⁴ X _c	⁰ X, ¹ X, ² X, ⁴ X _c , ⁵ X _{pc}	¹ X, ² X, ³ X _p , ⁴ X _p , ⁶ X _p
ST	⁰ X, ¹ X, ² X, ⁴ X _p , ⁵ X _p	¹ X, ² X, ⁴ X _p , ⁵ X _p , ⁶ X _p	⁰ X, ¹ X, ² X, ³ X _p , ³ X _c	¹ X, ⁴ X _p , ⁵ X _{pc} , ⁵ X _c , ⁶ X _{pc}
<i>V</i> _m	⁰ X, ¹ X, ² X, ³ X _p , ⁴ X _p	¹ X, ² X, ³ X _p , ⁴ X _{pc} , ⁵ X _c	⁰ X, ¹ X, ² X, ³ X _p , ³ X _c	¹ X, ² X, ³ X _p , ⁴ X _{pc} , ⁶ X _{pc}
<i>R</i> _m	⁰ X, ¹ X, ³ X _p , ⁴ X _p , ⁶ X _c	¹ X, ² X, ⁴ X _{pc} , ⁵ X _p , ⁶ X _c	⁰ X, ¹ X, ² X, ³ X _p , ⁴ X _c	⁰ X, ¹ X, ² X, ³ X _p , ⁴ X _{pc}
<i>V</i> _c	¹ X, ³ X _p , ³ X _c , ⁴ X _{pc} , ⁴ X _c	² X, ³ X _p , ³ X _c , ⁴ X _c , ⁶ X _c	⁰ X, ⁴ X _p , ⁴ X _{pc} , ⁶ X _p , ⁶ X _c	⁰ X, ¹ X, ² X, ³ X _p , ⁴ X _p
<i>P</i> _c	⁰ X, ³ X _c , ⁴ X _c , ⁴ X _{pc} , ⁶ X _c	¹ X, ² X, ⁴ X _{pc} , ⁶ X _c	⁰ X, ¹ X, ² X, ⁵ X _p , ⁶ X _c	¹ X, ² X, ³ X _p , ⁴ X _{pc} , ⁶ X _{pc}
<i>T</i> _c	⁰ X, ³ X _c , ⁴ X _p , ⁴ X _c , ⁵ X _{pc}	¹ X, ³ X _c , ⁴ X _{pc} , ⁴ X _c , ⁶ X _c	⁰ X, ¹ X, ² X, ⁴ X _p , ⁶ X _c	⁰ X, ¹ X, ² X, ³ X _p , ⁴ X _p

^a The overall connectivity (TC + TC1 + *K*) and molecular connectivity (χ) parameters included in the best five-parameter models are given in Table 2.

The major conclusion from Tables 5 and 6 is that the inverse-square-root connectivity function, a major component of molecular connectivity paradigm, does not show any advantage in linear structure–property relationships with the 10 examined physicochemical properties of C3–C8 alkanes. The performance of this function is about the same as that of the simple inverse function used in the ON set of indices, and slightly worse than that of the overall second Zagreb index OM2, which uses a direct but not inverse proportionality. However, all three connectivity functions employed in OM2-, ON-, and χ -type indices, and based on products of vertex degrees, perform considerably worse than overall connectivity indices TC, TC1, and their combination, all of them based on sums of vertex degrees. These indices produce the smallest standard deviations for nine of the examined 10 properties, and only for critical volume they fail in favor of the overall Zagreb index OM2. The physical meaning of this rather unexpected dominance of overall connectivity indices is still under investigation. It has to be emphasized that the overall connectivity and molecular connectivity indices assess the contributions coming from the subgraphs of certain size to be about the same. Thus, in both series of 50 best models compared in Table 2, about 40% of the variables incorporated are the three initial (zero-, first-, and second-order) terms. Another 20% comes from the two third-order terms in both series. Therefore, the difference in the performance of molecular connectivity and overall connectivity stems from the manner the contribution of these terms is defined — by product or by sums of vertex degrees, respectively. Having in mind that molecular connectivity indices have some quantum chemical basis [43,44], related to their “bond”-type description of molecular structure, one might expect quantum chemical arguments in favor of the overall connectivity to emerge from atoms-in-molecules types of theories. The excellent result obtained for the heat of atomization seems to support such expectations. Another argument might be inferred from the superposition principle in quantum mechanics, which for example requires bond wave functions to be formed from atomic wave functions by summation (superposition) but not multiplication. The physical interpretation of topological indices is generally an open question. More attention to this important topic will be devoted in a future publication [45].

5. Conclusion

This study demonstrates the usefulness of overall connectivity concept, which combines the basic ideas of the classical concept of molecular connectivity with those of molecular complexity. Within our concept, it is conjectured that an optimal description of molecular structure could be attained with topological indices, which match closely molecular complexity. More specifically, this implies descriptors that increase with all elements of molecular complexity, and first of all, with the increase of molecule size

and with the enhancing of typical complexity patterns like branching, cyclicity, and centrality. The inverse-square-root function used in molecular connectivity approach does not satisfy this requirement because the indices it defines increase with size but decrease with complexity. This is the first, but not the major reason for overall connectivity indices to produce multilinear models with lower standard deviation than those of molecular connectivity. By including the newly defined overall second Zagreb index [32,40,41] in the comparative study, we have an analog of molecular connectivity function, which satisfy the requirement for a concerted increase with size and complexity patterns. Yet, the OM2 index produced the best model for only one of the examined 10 physicochemical properties. For the other nine properties these were overall connectivity-type models. In the search for an explanation for this dominance, perhaps one should focus on the other major difference of overall versus molecular connectivity functions, namely the use of sums rather than products of vertex degrees in all connected subgraphs. This might more generally mean that an “atoms-in-molecule”-type of description could be advantageous in relating molecular structure to its properties and biological activities, as compared to “bonds-in-molecule” descriptions. Indeed, more investigations are needed to arrive at more definite explanation. Some of the properties examined may depend nonlinearly on topological descriptors (though hardly all 10 properties could behave that way), and a comparison of nonlinear models is needed to prove the dominance of overall connectivity indices. Whatever the final result might be, molecular connectivity concept’s 25th anniversary is the right time to revisit the entire paradigm of this powerful concept, which influenced greatly the search for new drugs and chemical products in the entire last quarter of the twentieth century.

Acknowledgements

The constructive proposals of the referees are gratefully acknowledged.

References

- [1] F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1969.
- [2] D. Bonchev, Information Theoretic Indices for Characterization of Chemical Structures, Research Studies Press, Chichester, UK, 1983.
- [3] I. Gutman, B. Rušćić, N. Trinajstić, C.W. Wilcox Jr., Graph theory and molecular orbitals. Part 12. Acyclic polyenes, J. Chem. Phys. 62 (1975) 3399–3405.
- [4] M. Randić, On characterization of molecular branching, J. Am. Chem. Soc. 97 (1975) 6609–6615.
- [5] L.B. Kier, L.H. Hall, W.J. Murray, M. Randić, Molecular connectivity. Part I. Relationship to nonspecific local anesthesia, J. Pharm. Sci. 64 (1975) 1971–1974.
- [6] G.A. Smolenskii, Zh. Fiz. Khim. 38 (1964) 1288.
- [7] M. Gordon, J.W. Kennedy, The graph-like state of matter. Part 2. TCGI schemes for the thermodynamics of alkanes and the theory

- of inductive inference, *J. Chem. Soc., Faraday Trans. II* 69 (1973) 484–504.
- [8] T. McLeish, On the trail of topological fluids, *Phys. World* 8 (1995) 32–37.
- [9] L.B. Kier, L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- [10] L.B. Kier, L.H. Hall, Derivation and significance of valence molecular connectivity, *J. Pharm. Sci.* 70 (1981) 583–589.
- [11] L.B. Kier, L.H. Hall, J.W. Frazer, An index of electrotopological state for atoms in molecules, *J. Math. Chem.* 7 (1991) 229–241.
- [12] A.T. Balaban, Highly discriminating distance-based topological index, *Chem. Phys. Lett.* 89 (1982) 399–404.
- [13] E. Estrada, Graph theoretical invariant of Randić revisited, *J. Chem. Inform. Comput. Sci.* 35 (1995) 1022–1025.
- [14] E. Estrada, N. Guevara, I. Gutman, Extension of edge connectivity index. Relationship to line graph indices and QSPR applications, *J. Chem. Inform. Comput. Sci.* 38 (1998) 428–431.
- [15] M. Randić, Augmented valence — novel graph theoretical index with an apparent interpretation. *Int. J. Quant. Chem.*, in press.
- [16] C. Shannon, W. Weaver, *Mathematical Theory of Communications*, University of Illinois Press, Urbana, MI, 1949.
- [17] H. Kastler (Ed.), *Essays on the Use of Information Theory in Biology*, University of Illinois Press, Urbana, MI, 1953.
- [18] N. Rashevsky, Life, information theory, and topology, *Bull. Math. Biophys.* 17 (1955) 229–235.
- [19] A. Mowshovitz, Entropy and the complexity of graphs. Part 1. An index of the relative complexity of a graph, *Bull. Math. Biophys.* 30 (1968) 175–204.
- [20] D. Minoli, Combinatorial graph complexity, *Atti. Acad. Naz. Lincei Rend.* 59 (1976) 651–661.
- [21] S.H. Bertz, The first general index of molecular complexity, *J. Am. Chem. Soc.* 103 (1981) 3599–3601.
- [22] S.H. Bertz, A mathematical model of molecular complexity, in: R.B. King (Ed.), *Chemical Applications of Topology and Graph Theory*, Elsevier, Amsterdam, 1983, pp. 206–221.
- [23] D. Bonchev, O.E. Polansky, On the topological complexity of chemical systems, in: R.B. King, D.H. Rouvray (Eds.), *Graph Theory and Topology in Chemistry*, Elsevier, Amsterdam, 1987, pp. 126–158.
- [24] S. Bertz, W.C. Herndon, *Similarity of graphs and molecules, Artificial Intelligence Applications in Chemistry*, ACS, Washington, DC, 1986, pp. 169–175.
- [25] D. Bonchev, Kolmogorov's information, Shannon's entropy, and topological complexity of molecules, *Bulg. Chem. Commun.* 28 (1995) 567–582.
- [26] D. Bonchev, Novel indices for the topological complexity of molecules, *SAR QSAR Environ. Res.* 7 (1997) 23–43.
- [27] D. Bonchev, Overall connectivity and molecular complexity, in: J. Devillers, A.T. Balaban (Eds.), *Topological Indices and Related Descriptors*, Gordon and Breach, Reading, UK, 1999, pp. 361–401.
- [28] S.H. Bertz, T.J. Sommer, Rigorous mathematical approaches to strategic bonds and synthetic analysis based on conceptually simple new complexity indices, *Chem. Commun.* (1997) 2409–2410.
- [29] S.H. Bertz, W.F. Wright, The graph theory approach to synthetic analysis: definition and application of molecular complexity and synthetic complexity, *Graph Theory Notes N.Y. Acad. Sci.* 35 (1998) 32–48.
- [30] D. Bonchev, D.H. Rouvray (Eds.), *Complexity in Chemistry, Mathematical Chemistry Series, Vol. 7*, Gordon and Breach, Amsterdam, in press.
- [31] D. Bonchev, E. Gordeeva, Partial orderings induced by topological complexity, *MATCH* 42 (2000) 85–117.
- [32] S. Nikolić, N. Trinajstić, I.M. Tolić, G. Rücker, C. Rücker, On molecular complexity indices, in: *Complexity in Chemistry, Mathematical Chemistry Series, Vol. 7*, Gordon and Breach, Amsterdam, in press.
- [33] G. Rücker, C. Rücker, Walk counts, labyrinthicity, and complexity of acyclic and cyclic graphs and molecules, *J. Chem. Inform. Comput. Sci.* 40 (2000) 99–106.
- [34] D. Bonchev, Overall connectivities/topological complexities: a new powerful tool for QSPR/QSAR, *J. Chem. Inform. Comput. Sci.* 40 (2000) 934–941.
- [35] D.E. Needham, I.-C. Wei, P.G. Seybold, Molecular modeling of the physical properties of the alkanes, *J. Am. Chem. Soc.* 110 (1988) 4186–4194.
- [36] W.C. Herndon, 1997, private communication.
- [37] *Selected Values of Properties of Hydrocarbons and Related Compounds*, American Petroleum Institute, Research Project 44, 1977, Pittsburgh, Pennsylvania.
- [38] M. Garbalena, W.C. Herndon, Optimum graph-theoretical models for enthalpic properties of alkanes, *J. Chem. Inform. Comput. Sci.* 32 (1992) 37–42.
- [39] O. Mekenyan, S. Karabunarliev, D. Bonchev, The OASIS concept for predicting biological activity of chemical compounds, *J. Math. Chem.* 4 (1990) 207–215.
- [40] A. Golbraikh, D. Bonchev, A. Tropsha, Novel chirality descriptors derived from molecular topology, *J. Chem. Inform. Comput. Sci.* 41 (2001) 147–158.
- [41] D. Bonchev, N. Trinajstić, Overall connectivity indices. Part 3. Overall Zagreb indices, *SAR QSAR Environ. Res.* 12 (2001) 213–235.
- [42] O. Ivanciuc, 2000, private communication.
- [43] I.V. Stankevich, M.I. Skvortsova, N.S. Zefirov, On a quantum chemical interpretation of molecular connectivity indices for conjugated hydrocarbons, *J. Mol. Struct. (Theochem)* 342 (1995) 173–179.
- [44] J. Galvez, On a topological interpretation of electronic and vibrational molecular energies, *J. Mol. Struct. (Theochem)* 429 (1998) 255–264.
- [45] D. Bonchev, D.H. Rouvray (Eds.), *Topological Descriptors of Molecular Structure. Theory, Interpretation and Application*, Mathematical Chemistry Series, Vols. 8/9, Gordon and Breach, Amsterdam, in preparations.