Letter

GTI-space: the space of generalized topological indices

Ernesto Estrada*

Complex Systems Research Group, X-Rays Unit, RIAIDT, Edificio CACTUS,
University of Santiago de Compostela, Santiago de Compostela 15782, Spain
E-mail: estrada66@yahoo.com

Adelio R. Matamala

Facultad de Ciencias Químicas, Universidad de Concepción, Casilla 160-C, Concepción, Chile

A new extension of the generalized topological indices (GTI) approach is carried out to represent "simple" and "composite" topological indices (TIs) in an unified way. This approach defines a GTI-space from which both simple and composite TIs represent particular subspaces. Accordingly, simple TIs such as Wiener, Balaban, Zagreb, Harary and Randić connectivity indices are expressed by means of the same GTI representation introduced for composite TIs such as hyper-Wiener, molecular topological index (MTI), Gutman index and reverse MTI. Using GTI-space approach we easily identify mathematical relations between some composite and simple indices, such as the relationship between hyper-Wiener and Wiener index and the relation between MTI and first Zagreb index. The relation of the GTI-space with the sub-structural cluster expansion of property/activity is also analysed and some routes for the applications of this approach to QSPR/QSAR are also given.

KEY WORDS: graph theory, generalized graph matrix, topological indices, molecular descriptors, algebraic graph theory

1. Introduction

Topological molecular descriptors, the so-called topological indices (TIs), have proved to be of great usefulness and effectiveness in molecular design [1–3]. The main drawback of these descriptors is the lack of a unified way for their definition. As a matter of comparison all quantum atomic and molecular descriptors are defined on the basis of the molecular wavefunction. However, TIs are generally defined through the use of several invariants applied to different algebraic representation of molecular graphs [4]. The most used of such representations are the adjacency (**A**) and distance (**D**) matrices of the graph

*Corresponding author.

508

[4]. Despite the fact that these two matrices are not unrelated [5], the great variability and the *ad hoc* nature of the structural invariants used to define TIs gives the impression of a profound lack of unity among all these molecular descriptors.

In an attempt for unifying some of the best known TIs one of the current authors introduced the generalized graph matrix, Γ [6]. Using this matrix and an approach based on a vector–matrix–vector multiplication procedure [7, 8] it is possible to defined several of the "classical" TIs in an unified way with applications in structure-property relationships [9–11] and other branches of theoretical chemistry [12, 13]. According to this approach, the adjacency and distance matrices are particular cases of an infinite set of matrices generated from Γ [6]. At the same time several "classical" TIs, such as Wiener (W) index [14], Balaban J index [15], Harary H indices [16, 17], Zagreb M indices [18] and Randić (χ) connectivity index [19] are defined on the basis of the same graph invariant. Some of these indices, such as W, J, H_1 and H_2 are based on the distance matrix while the others (M_1, M_2 and χ) are based on the adjacency matrix of the molecular graph. Thus, we propose to designate these descriptors as "simple" TIs due to the fact that they are based on a single topological matrix. On the other hand, there are several TIs that should be designated as "composite" because they are based on invariants that use more than one single matrix in their definition. These are the cases of descriptors like the Gálvez charge indices [20], the Molecular Topological Index (MTI) [21] or the hyper-Wiener index [22]. In the first two examples the indices are defined by using both the adjacency and distance matrices, while the last one is defined on the basis of the distance matrix and a matrix whose elements are the squares of the distances between pairs of vertices in the graph. Another index was introduced by Gutman [23] in the context of the Schultz MTI [21] and has been designated as the Gutman molecular topological index in the Todeschini–Consonni book [24]. It is based on a combination of the adjacency and distance matrices and deserves the qualification of "composite" TI. Among these descriptors the Schultz MTI [21] and the hyper-Wiener indices [22] have received a great deal of attention in both mathematical and chemical literature [23, 25–29]. In the current work we propose to extend our Generalized topological indices (GTI) approach to include not only "simple" but also "composite" TIs, such as MTI and hyper-Wiener index. In doing such extension we define the GTI-space, from which simple and composite GTIs are particular subspaces. According to this approach new GTIs can be obtained by linear combinations of other GTIs, which can explain the relationships previously found by others between the composite MTI and hyper-Wiener with the simple Wiener index.

2.   Generalized topological indices

Let $G(V, E)$ be a molecular-graph with $|V| = n$ vertices and $|E| = m$ edges. Let d_{ij} be the entries of the $n \times n$ topological distance matrix of the graph $G(V, E)$. The GTI associated to the graph $G(V, E)$ is defined by the following vector–matrix–vector formula:

$$\text{GTI}[G] = C \left(\begin{array}{ccc|c} p & q & r & \bar{s} \\ x & y & z & \bar{w} \end{array} \right)_G = C\, \mathbf{u}^{\mathrm{T}}(G; y, \overline{w}, q)\Gamma(G; x, p)\, \mathbf{v}(G; z, \bar{s}, r), \quad (1)$$

where C is a constant and u and v are column vectors whose components are given by

$$u_i(G; y, q) = \left[w_i + \sum_{j=1}^{n} g_{ij}(y, 1) \right]^q \quad \text{and} \quad v_i(G; z, r) = \left[s_i + \sum_{j=1}^{n} g_{ij}(z, 1) \right]^r. \quad (2)$$

The Γ matrix is the so-called *generalized molecular-graph matrix* whose $n \times n$ entries are expressed in terms of the topological distance through

$$g_{ij}(G; x, p) = \begin{cases} 1, & \text{if } d_{ij} = 1, \\ \left(d_{ij} x^{d_{ij}-1} \right)^p, & \text{if } i \neq j; d_{ij} \neq 1, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The vectors $\overline{w} = (w_1, w_2, \ldots, w_n)$ and $\bar{s} = (s_1, s_2, \ldots, s_n)$ contain the weighting parameters for differentiating heteroatoms in certain TIs, such as the valence connectivity index. In general, for TIs $\overline{w} = \bar{s} = (0, 0, \ldots, 0) = \bar{0}$.

From equation (1), each GTI can be written in the following form:

$$\text{GTI}[G] = C \sum_{k=1}^{\text{diam}(G)} c_k\, \eta^{(k)}[G], \quad (4)$$

where diam(G)= max (k) is the diameter of the graph G, i.e. the largest geodesic (k) in the graph G,

$$c_k = c_k(x, p) = k^p\, x^{p(k-1)} \quad (5)$$

and

$$\eta^{(k)}[G] = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \langle i, j \rangle_G\, \Delta_{ij}^{(k)}. \quad (6)$$

The kth order geodesic (shortest path) matrix $\Delta^{(k)}$ of the graph G is defined by the following entries:

$$\Delta_{ij}^{(k)}[G] = \begin{cases} 0, & \text{if } d_{ij} \neq k \text{ in graph } G, \\ 1, & \text{if } d_{ij} = k \text{ in graph } G. \end{cases} \quad (7)$$

And, each geodesic-bracket $\langle i, j \rangle_G$ is defined by

$$\langle i, j \rangle_G = \frac{1}{2} \left(u_i v_j + v_i u_j \right). \tag{8}$$

Functions $u_i(G; y, q)$ and $v_i(G; z, r)$ adopts the form:

$$u_i(G; y, q, \overline{w}) = \left[w_i [G] + \delta_i [G] + \sum_{k=2}^{\mathrm{diam}(G)} k \, y^{k-1} N_i^{(k)} [G] \right]^q, \tag{9}$$

$$v_i(G; z, r, \overline{s}) = \left[s_i [G] + \delta_i [G] + \sum_{k=2}^{\mathrm{diam}(G)} k \, z^{k-1} N_i^{(k)} [G] \right]^r, \tag{10}$$

where $\delta_i [G]$ is the classic degree of the ith vertex in the graph G, whereas the quantity $N_i^{(k)} [G]$ is the number of vertices at distance k from the ith vertex:

$$N_i^{(k)} [G] = \sum_{j=1}^{n} \Delta_{ij}^{(k)} [G]. \tag{11}$$

Equation (4) shows that any GTI can be separated in terms of the contributions of pair of vertices at the same topological distance in the graph. Each $\eta^{(k)}$ term defines the contribution to GTI of all *interactions* between pairs of vertices separated at distance k in the graph. These contributions are scaled by two real parameters through the c_k coefficients. The diameter of the graph is a global descriptor and its presence in definition (4) reveals the dependence of any GTI on the "size" of the molecular-graph. On the other hand, the molecular-connectivity relationship among atoms in the molecule defines the "shape" of the molecular graph. This "shape" is coded by the so-called *geodesic-brackets* (see equation 8). Equations (9) and (10) show that the functions u and v are the generalization of the "classic" vertex degree notion. Through these functions u and v and by settling the x and p parameters, a pair of weights is assigned to each vertex in the graph. On each vertex these weights code the topological environment around it. From the previous analysis, it is clear that the codification of the topological complexity of any molecular-graph relies on the $N_i^{(k)} [G]$ quantities, defined by equation (11).

3.   GTI-space and "simple" TIs

The set of all GTI forms a non-linear real space of functions depending on $(2n + 6)$ parameters: the scalars x, y, z, p, q, r and components of $\overline{w} = (w_1, w_2, \ldots, w_n)$ and $\overline{s} = (s_1, s_2, \ldots, s_n)$ vectors. This space will be termed

the GTI-space. An important subspace of this space is formed by the "classic" topological indices. From (1), it is straightforward to obtain several of the well-known classical indices. For instance, the Zagreb indices, $M_1(G)$ and $M_2(G)$; the Randić connectivity index, $\chi(G)$; the Wiener index, $W(G)$, the Balaban $J(G)$ and the Harary numbers $H_1(G)$ and $H_2(G)$, which will be designated as $H_k(G)$, are expressed as follow:

$$M_1[G] = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_G, \tag{12}$$

$$M_2[G] = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}_G, \tag{13}$$

$$\chi[G] = \frac{1}{2} \begin{pmatrix} 1 & -1/2 & -1/2 \\ 0 & 0 & 0 \end{pmatrix}_G, \tag{14}$$

$$W[G] = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}_G, \tag{15}$$

$$J[G] = \frac{1}{2B} \begin{pmatrix} 1 & -1/2 & -1/2 \\ 0 & 1 & 1 \end{pmatrix}_G, \quad \text{where } B = m/(m-n+2), \tag{16}$$

$$H_k[G] = \frac{1}{2} \begin{pmatrix} k & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}_G. \tag{17}$$

For the sake of simplicity we have used the following compact symbol when $\overline{w} = (0, 0, \ldots, 0) = \overline{0}$ and $\overline{s} = (0, 0, \ldots, 0) = \overline{0}$:

$$\begin{pmatrix} p & q & r \\ x & y & z \end{pmatrix} \begin{Vmatrix} \overline{0} \\ \overline{0} \end{Vmatrix}_G = \begin{pmatrix} p & q & r \\ x & y & z \end{pmatrix}_G. \tag{18}$$

4.   GTI-space and "composite" TIs

The first composite index that we will analyse here is the one introduced by Gutman in the context of the MTI [23]. It was originally defined on the basis of the adjacency and distance matrices as follows:

$$S_G(G) = \sum_i \sum_j (ADA)_{ij} = \sum_i \sum_j \delta_i \delta_j d_{ij}. \tag{19}$$

This index can be simply written in the context of GTI as follows:

$$S_G[G] = \begin{pmatrix} 1\ 1\ 1 \\ 1\ 0\ 0 \end{pmatrix}_G. \tag{20}$$

On the other hand, the Schultz MTI (G) [21] was originally defined as follows:

$$\text{MTI}(G) = \sum_i [(\mathbf{A} + \mathbf{D})\,\mathbf{v}]_i, \tag{21}$$

where $\mathbf{v}$ is the column vector of vertex degrees. MTI can be defined in the context of the GTI approach as follows:

$$\text{MTI}(G) = \begin{pmatrix} 1\ 1\ 0 \\ 0\ 0\ 0 \end{pmatrix}_G + \begin{pmatrix} 1\ 1\ 0 \\ 1\ 0\ 0 \end{pmatrix}_G. \tag{22}$$

Schultz and Schultz defined a reciprocal MTI, which is based on the matrix of reciprocal topological distances $\mathbf{RD}$ [30]:

$$\text{RMTI}(G) = \sum_i [(\mathbf{A} + \mathbf{RD})\,\mathbf{v}]_i. \tag{23}$$

This index is expressed as follows in the GTI context:

$$\text{RMTI}(G) = \begin{pmatrix} 1\ 1\ 0 \\ 0\ 0\ 1 \end{pmatrix}_G + \begin{pmatrix} -1\ 1\ 0 \\ 1\ \ 0\ 1 \end{pmatrix}_G. \tag{24}$$

Finally, we will consider the hyper-Wiener index introduced by Randić [22] and generalized by Klein et al. [25]. According to the general definition of this index [25]:

$$\text{WW}(G) = \frac{1}{4} \sum_i \sum_j \left(d_{ij}^2 + d_{ij} \right) \tag{25}$$

Thus, it is easy to see that it can be expressed using the generalized graph matrix as follows:

$$\text{WW}(G) = \frac{1}{4} \begin{pmatrix} 1\ 0\ 0 \\ 1\ 1\ 1 \end{pmatrix}_G + \frac{1}{4} \begin{pmatrix} 2\ 0\ 0 \\ 1\ 1\ 1 \end{pmatrix}_G. \tag{26}$$

In this section, we have been able to express several composite TIs in terms of the GTI approach. Accordingly, all these composite TIs are not only expressed in an unified way but also they are expressed using the same algebraic expressions as for the simple TIs, which indicates a direct route for the generalization of both types of descriptors.

5.   Generalization of "simple" and "composite" TIs

From the previous results, we postulate a new class of GTI defined by:

$$\text{GTI}[G] = C\mathbf{u}^{\text{T}}(G; y, \overline{w}, q) \left[\sum_{i=1} \gamma_i' \Gamma(G; x_i, p_i) \right] \mathbf{v}(G; z, \overline{s}, r), \tag{27}$$

where γ_i' is a coefficient modifying the generalized graph matrices. Thus, for a given set of parameters *y, z, q, r* and components of $\overline{w} = (w_1, w_2, \ldots, w_n)$ and $\overline{s} = (s_1, s_2, \ldots, s_n)$ vectors, we obtain

$$\text{GTI}[G] = \sum_{i=1} \gamma_i \left(\mathbf{u}^{\text{T}}(G; y, \overline{w}, q) \Gamma(G; x_i, p_i) \, \mathbf{v}(G; z, \overline{s}, r) \right), \tag{28}$$

where γ_i coefficients contain the information of both C and γ_i'. Equation. (28) can be written in the equivalent form:

$$\text{GTI}[G] = \sum_{i=1} \gamma_i \, \text{GTI}_i[G], \tag{29}$$

where

$$\text{GTI}_i[G] = \mathbf{u}^{\text{T}}(G; y, \overline{w}, q) \Gamma(G; x_i, p_i) \, \mathbf{v}(G; z, \overline{s}, r). \tag{30}$$

Equation (28) defines a linear subspace of GTIs. Accordingly, these "composite" TIs are linear combinations of other "simple" descriptors. Lets take, for instance, the expression (22) defining MTI in the context of the GTI-space. It is evident that the first bracket of this expression corresponds to the expression of the Zagreb M_1 index given by expression (12). The second bracket can be expressed as follows:

$$S(G) = \begin{pmatrix} 1 \ 1 \ 0 \\ 1 \ 0 \ 0 \end{pmatrix}_G = \mathbf{v}\mathbf{D}\mathbf{u}^{\text{T}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i d_{ij}. \tag{31}$$

This index was originally defined by Müller et al. [31] and by Mihalic et al. [32] and called MTI' index. Thus, using Zagreb M_1 and $S(G)$ we arrive to the conclusion that MTI $= M_1 + S$, which was previously derived by Gutman [23] (see that in Gutman paper [23] as well as in Todeschini–Connsoni book [24] the M_1 is designated as M_2, which can produce some confusion with the second Zagreb index).

In the case of the hyper-Wiener index it is easy to identify the first bracket as the Wiener index and the second one as the sum of the squared distances in the graph, which is known as the unnormalized second moment of distances, D_2. Then, it is obvious that WW $= 1/2\,(W + D_2)$ as previously recognized by Klein et al. [25]. Thus, it is evident from the previous results that the GTI-space

approach permits to find mathematical relations between TIs. Of course, the further exploration of other GTI-space properties (metric, topological, etc.) will make possible the discovery of new and unnoticed relationships between these descriptors.

6.   GTI-space: some routes to applications

One of the most important advantages that the use of the GTI offers to develop QSAR/QSPR models is the possibility of optimizing TIs to describe a particular property, P [9–12]. Saying it in other words, TIs are *ad hoc* molecular descriptors which are not optimal for describing a particular property/activity. They correspond to an initial "configuration" of the GTI parameter space that need to be optimized to describe P in an optimal way. A full analogy can be found with the use of quantum chemical molecular approaches where an optimal geometry is searched for minimizing the energy of the molecule. Here, the initial atomic coordinates are the non-optimal set of parameters which are optimized to minimize the molecular energy. Thus, the optimization of the GTIs consists in finding the best set of parameters that minimize the error in predicting a property/activity under study. For the sake of simplicity we consider here a linear model relating the property P and the GTIs:

$$P[G] = \sum_{i=1} \gamma_i \, \mathrm{GTI}_i[G] + \text{Error}. \tag{32}$$

Here, we can consider the coefficients γ_i of the GTI as the coefficients of the regression model, which means that the linear combination of GTIs obtained for describing a specific property/activity is also a GTI:

$$P[G] - \text{Error} = \mathrm{GTI}[G] = \sum_{i=1} \gamma_i \, \mathrm{GTI}_i[G]. \tag{33}$$

This approach, which considers a "topological function" as a new TI, has been previously used in the literature, in particular by considering a linear discriminant function based on TIs as a "super-TI" to discriminate/predict other biological activities [33].

On the other hand, we have previously shown that the GTIs can be expressed as linear combinations of substructures ξ of the graph G, $f(G, \xi)$, which permits to write the expression (32) as a *sub-structural cluster expansion* of the form [34, 35]:

$$P[G] = \sum_{\xi}^{\subseteq G} P_f(\xi) \, f(G, \xi), \tag{34}$$

where the parametric coefficients $P_f(\xi)$ depend exclusively on the GTI coefficients γ_i and $f(G, \xi)$ depends solely on the way in which the sub-structures embed in the graph.

7.   Concluding remarks

One of the most important targets in science is the development of theories and approaches explaining (apparently) diverse phenomena in an unified way. The best known example is the *unified field theory*, which is an attempt to unify all the fundamental forces and the interactions between elementary particles into a single theoretical framework. In a much modest level the current approach pretends to represent several disparate mathematical representations of a discrete object in an unified formulation. There is both theoretical and practical appeal for atempting the unification of TIs under a common umbrella of a generalized graph approach. On the first side, the generalized approach represents a condensation of great part of the knowledge about topological molecular invariants in a formal way. This permits to understand the nature of these descriptors, their interrelationships and structural interpretation in a better way than by studying TIs on a one-by-one basis. On the practical side, the GTI-space permits the optimization of the TIs to describe a property/activity in a most efficient way than the unoptimized descriptors which have been introduced in an *ad hoc* way. The facilities offered by the current approach to identify mathematical relationships between TIs, which are further translated into intercorrelations between such indices, permit to avoid the unnecessary proliferation of "new" topological descriptors. We hope that the further exploration of algebraic, metric and topological properties of GTI-space will open new avenues in chemical graph theory.

Acknowledgments

References

[1] J. Devillers and A.T. Balaban (eds.), *Topological Indices and Related descriptors in QSAR and QSPR* (Gordon & Breach, Amsterdam, 1999).
[2] E. Estrada and E. Uriarte, Curr. Med. Chem. 8 (2001) 1573–1588.
[3] J.R. Votano, Curr. Opin. Drug Discov. Dev. 8 (2005) 32–37.
[4] N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, FL, 1992).
[5] R.F. Brown and B.W. Brown, *Finite Mathematics* (Ardsley, New York, 1992) pp. 534–545.
[6] E. Estrada, Chem. Phys. Lett. 336. (2001) 248–252.
[7] E. Estrada, L. Rodríguez and A. Gutiérrez, MATCH: Comm. Math. Comput. Chem. 35 (1997) 145–156.
[8] E. Estrada and L. Rodríguez, MATCH: Comm. Math. Comput. Chem. 35 (1997) 157–167.

 [9] E. Estrada and Y. Gutierrez, MATCH: Comm. Math. Comput. Chem. 44 (2001) 155–167.
[10] A.R. Matamala and E. Estrada, Chem. Phys. Lett. 410 (2005) 343–347.
[11] A.R. Matamala and E. Estrada, J. Phys. Chem. A 109 (2005) 9890–9895.
[12] E. Estrada, J. Phys. Chem. A 107 (2003) 7482–7489.
[13] E. Estrada, J. Phys. Chem. A 108 (2004) 5468–5473.
[14] H. Wiener, J. Am. Chem. Soc. 69 (1947) 17–20.
[15] A.T. Balaban, Chem. Phys. Lett. 89 (1982) 399–404.
[16] D. Plavšić, S. Nikolić, N. Trinajstić and Z. Mihalić, J. Math. Chem. 12 (1993) 235–250.
[17] O. Ivanciuc, T.S. Balaban and A.T. Balaban, J. Math. Chem. 12 (1993) 309–318.
[18] I. Gutman, B. Ruscić, N. Trinajstić and C.F. Wilcox, J. Chem. Phys. 62 (1975) 3399–3405.
[19] M. Randić, J. Am. Chem. Soc. 97 (1975) 6609–6615.
[20] J. Gálvez, R. García-Domenech, M.T. Salabert and R. Soler, J. Chem. Inf. Comput. Sci. 34 (1994) 520–525.
[21] H.P. Schultz, J. Chem. Inf. Comput. Sci. 29 (1989) 227–228.
[22] M. Randić, Chem. Phys. Lett. 211 (1993) 478–483.
[23] I. Gutman, J. Chem. Inf. Comput. Sci. 34 (1994) 1087–1089.
[24] R. Todeschini and V. Consoni, Handbook of Moelcular Descriptors (Wiley-VCH, Weinheim, 2000).
[25] D.J. Klein, I. Lukovits and I. Gutman, J. Chem. Inf. Comput. Sci. 35 (1995) 50–52.
[26] I. Gutman, Chem. Phys. Lett. 364 (2002) 352–356.
[27] I. Gutman, J. Serbe. Chem. Soc. 68 (2003) 949–952.
[28] G.G. Cash, Appl. Math. Lett. 15 (2002) 893–895.
[29] B. Zhou and I. Gutman, Chem. Phys. Lett. 394 (2005) 93–95.
[30] H.P. Schultz and T.P. Schultz, J. Chem. Inf. Comput. Sci. 38 (1998) 853–857
[31] W.R. Müller, K. Szymanski, J.V. Knop and N. Trinajstić, J. Chem. Inf. Comput. Sci. 30 (1990) 160–163.
[32] Z. Mihalić, S. Nikolić and N. Trinajstić, J. Chem. Inf. Comput. Sci. 32. (1992) 28–37.
[33] J. Gálvez, R. García-Domenech, V. De Julian-Ortiz and R. Soler, J. Chem. Inf. Comput. Sci. 34 (1994) 1198–1203.
[34] D.J. Klein, Int. J. Quantum Chem. S20 (1986) 153–171.
[35] D.J. Klein, T.G. Schmaltz and L. Bytautas, SAR QSAR Environ. Res. 10 (1990) 131–156.