

STATISTICAL-MECHANICAL THEORY OF TOPOLOGICAL INDICES

ERNESTO ESTRADA

ABSTRACT. Topological indices (TI) are algebraic invariants of molecular graphs representing the topology of a molecule, which are very valuable in quantitative structure-property relations (QSPR). Here we prove that TI are the partition functions of such molecules when the temperature of the thermal bath at which they are submerged is very high. These partition functions are obtained by describing molecular electronic properties through tight-binding Hamiltonians (TBH), where the hopping parameters are topological properties describing atom-atom interactions. We prove that the TBH proposed here are non-Hermitian diagonalizable Hamiltonians which can be replaced by symmetric ones. In this way we propose a statistical-mechanical theory for TI, which is exemplified by deriving the Randić, Zagreb, Balaban, Wiener and ABC indices. The work also illuminates how to improve QSPR models using the current theoretical framework as well as how to derive statistical-mechanical parameters of molecular graphs.

Keywords: topological indices, partition function, tight-binding Hamiltonian, graph theory, statistical mechanics

1. INTRODUCTION

Nowadays the use of networks to represent the skeleton of complex systems is a well-established approach in physical sciences [1–3]. A network is a graph $G = (V, E)$ with $\#V = n$ nodes (vertices) representing the entities of the system and $\#E = m$ edges representing the interactions between pairs of entities [2]. In this context the use of graphs to represent molecules has a long history dating back to the development of graph theory as a discipline [4]. A molecular graph is a graph in which the nodes represent atoms and the edges represent covalent bonds between pairs of atoms in a molecule [5, 6]. A way of characterizing the structure of such molecular graphs is by means of graph-theoretic invariants known as topological indices (TI) [7, 8]. One of the most successful TI is the so-called Randić index χ [9]. If we denote by k_v the degree of the node v , i.e., the number of connections that v has, the Randić index is defined as:

$$(1.1) \quad \chi := \sum_{(v,w) \in E} (k_v k_w)^{-1/2}.$$

This index, proposed by Milan Randić in 1975 [10], has been widely applied to the quantitative correlation with molecular properties [11, 12], which is known as QSPR studies (QSPR is the acronym for Quantitative Structure-Property Relations). In general, other TI can be defined using a similar type of graph-theoretic invariant. That is, if e_{vw} is an edge with endpoints v and w , we can define a few TI of the form:

$$(1.2) \quad \mathcal{I}(G) := \sum_{(v,w) \in E} f(e_{vw}),$$

where $f(e_{vw})$ is a function of the edge.

The main problem with the use of TI in QSPR is not, in general, of a quantitative nature. That is, either by using simple or sophisticated chemometric tools, TI have proved to be appropriate for describing quantitatively many physico-chemical, biological, environmental and toxicological properties of organic compounds [13–16]. The problem is then of a qualitative nature. In contrast with other molecular descriptors which are defined on the basis of geometric parameters of molecules or from well-established quantum-mechanical principles [17], TI are based on an ad-hoc description of a molecule as a graph. The same criticism could be levied against the use of graphs to represent an air transportation system [18], the neural connections in the brain [19], or a social group [20]. But, while in these previously mentioned systems the notion of transport is well-understood at a topological level, the use of molecular graphs has been more fiercely criticized (see [21] and references therein), possibly due to the existence of an alternative “first-principles” way of representing molecules by using quantum-mechanical principles.

Although some attempts have been made to explain what a TI is after all, none of them can be considered as a physical derivation of these indices [22–26]. We can identify two kinds of approaches in these studies. From one side, the use of chemometric techniques to analyze these descriptors. On the other hand, the use of mathematical, mainly algebraic, approaches to understand the nature of the invariants in which TI are based on. An example of the first group is the paper [22] entitled “The chemical meaning of topological indices” published in 1992 by Todeschini et al. This paper gives a taxonomic classification of TI in groups obtained using principal component analysis. On the second group we can mention the work [23] “Generalization of topological indices” where it is shown that several TI are based on a generalized quadratic form of graph-theoretic matrices. Such taxonomies are important to understand the classes of TI existing but they do not answer the question of what TI are.

Here, we approach the answer to the question of what is a TI from first-principles physical concepts. First, we consider the description of molecules by using tight-binding Hamiltonians (TBH). The hopping parameter of these TBH are defined from the basis of the connectivity properties of atoms in a molecule. Then, we consider that such molecules are submerged into thermal baths of given temperature and apply Boltzmann formalism of statistical mechanics to obtain the probability of finding a molecule in a state of a given energy. Such energy is described by the eigenvalues of the TBH as usual. We then prove that the TBH proposed here are non-Hermitian diagonalizable Hamiltonians which can be replaced by symmetric ones. More importantly, we prove that TI are obtained as the partition function of the molecule in the corresponding TBH formalism when the temperature is very high. This derivation is detailed for the case of the Randić index and then extended to Zagreb indices, the Balaban index, and the ABC index, among others. The current first-principles formalism not only answers the question: what is a TI? from physical grounds but also gives the possibility of improving the quality of QSPR models using them as illustrated for the case of the Randić index.

2. STATISTICAL MECHANICS DERIVATION OF THE RANDIĆ INDEX

Let us consider the molecular graph $G = (V, E)$ representing a given molecule \mathcal{M} . Here we describe the electronic structure of the molecule \mathcal{M} by means of the TBH of the form [27]:

$$(2.1) \quad \mathcal{H} := \sum_i \alpha_{ii} c_i^\dagger c_i + \sum_{ij} t_{ij} c_i^\dagger c_j A_{ij},$$

where $c_i^\dagger (c_i)$ is the annihilation (creation) operator of a particle at the atom i of G , α_{ii} is the attraction force of the particle by the node i , and t_{ij} is the transfer energy of the particle from atom i to j , also known as the hopping parameter. The use of a TBH to describe saturated hydrocarbons, where there are strong covalent two-centered directional bonds, is not infrequent. For instance, Yang and Mak [28] used a TBH to describe these molecular systems by selecting four types of hopping parameters for carbon-carbon interactions and two for carbon-hydrogens. In this way they described very well properties, such as bond length and angles, vibrational frequencies, bond energies, and atomization energies, of alkanes as well as unsaturated hydrocarbons. Pan [29] used a similar TBH with different parameterization to describe geometric parameters and atomization energies of saturated and unsaturated hydrocarbons. Another TBH was developed by Horsfield et al. [30] for describing geometric and energetic properties of hydrocarbons, including linear alkanes. Here we will consider a variation of these approaches in which we will use topological parameters for the hopping parameter instead of distance-dependent parameterizations as used in the previously mentioned works. An advantage of the TBH is that because there is no electron-electron interaction, the Hamiltonian can be exactly diagonalized as a noninteracting system. Here we will construct TBH which are not symmetric, but which are similar to symmetric ones, such that they can always be exactly diagonalized.

Hereafter, we will consider $\alpha_{ii} = 0$ for all nodes $i \in V$. Now we introduce the following criterion for the inter-nodal transfer energy. Here we consider that a particle located at the node $i \in V$ is attracted by the node $j \in V$ with a force which is inversely proportional to the square root of the degree of the node i , k_i , such that

$$(2.2) \quad \mathcal{H}_R := - \sum_{ij} k_i^{-1/2} c_i^\dagger c_j A_{ij},$$

which means that

$$(2.3) \quad \mathcal{H}_R := -K^{-1/2}A,$$

where K is the diagonal matrix of node degrees. This TBH is non-symmetric in general because $\mathcal{H}_{R,ij} = k_i^{-1/2}$ and $\mathcal{H}_{R,ji} = k_j^{-1/2}$. However, we have the following result.

Lemma 1. *Let $R := K^{-1/2}A$, where K is the diagonal matrix of node degrees. Then,*

(i) the eigenvalues of R are real;

(ii) R can be diagonalized as $R = (K^{-1/4}J) \Xi (K^{-1/4}J)^{-1}$ where $K^{-1/4}AK^{-1/4} = J \Xi J^T$.

Proof. We start by showing that

$$(2.4) \quad R = K^{-1/2}A = K^{-1/4} \left(K^{-1/4}A \right) = K^{-1/4} \left(K^{-1/4}AK^{-1/4} \right) K^{1/4},$$

thus, R is similar to the symmetric matrix $(K^{-1/4}AK^{-1/4})$, and so its eigenvalues are real. Then we have that because $(K^{-1/4}AK^{-1/4})$ is symmetric we can write it as: $K^{-1/4}AK^{-1/4} = J \Xi J^{-1}$. Thus, $R = (K^{-1/4}J) \Xi (K^{-1/4}J)^{-1}$, which indicates that R is diagonalizable. \square

The previous result indicates that \mathcal{H}_R is a non-Hermitian diagonalizable TBH and so it can be replaced by a Hermitian Hamiltonian which is similar to it. Let us consider that the molecule \mathcal{M} is submerged into a thermal bath of inverse temperature $\beta = (k_B T)$ where k_B is the Boltzmann constant and T is the temperature [31]. Here, the thermal bath represents the following. Let N be the total number of molecules of a given substance. Let \mathcal{M} be a specific molecule of this substance for which we are writing the corresponding TBH. Then, the thermal bath is formed by the $N - \mathcal{M}$ molecules in which \mathcal{M} is submerged to. Therefore, according to the Boltzmann distribution law, the probability P_j of finding the molecule \mathcal{M} in the particular state j with energy E_j is

$$(2.5) \quad P_j = \frac{e^{-\beta E_j}}{\sum_j e^{-\beta E_j}} = \frac{e^{\beta \sigma_j}}{\sum_j e^{\beta \sigma_j}},$$

where σ_j is the j th eigenvalue of $R = K^{-1/2}A$. Thus, the term

$$(2.6) \quad Z := \sum_j e^{\beta \sigma_j} = \text{tr} (e^{\beta R}),$$

is the partition function of the molecule in the current theoretical framework. We then have the following result.

Lemma 2. *Let $0 < \epsilon \ll 1$ be a constant. Then, the Randić index is proportional to the partition function of the molecular graph when $\beta \rightarrow \epsilon$:*

$$(2.7) \quad \chi(G) = \lim_{\beta \rightarrow \epsilon} \frac{1}{\beta^2} (Z - n).$$

Proof. Let us start by writing the power-series expansion of the partition function, such that

$$(2.8) \quad Z_R(\beta) = \text{tr} \frac{(\beta R)^0}{0!} + \text{tr} \frac{(\beta R)}{1!} + \text{tr} \frac{(\beta R)^2}{2!} + \text{tr} \frac{(\beta R)^3}{3!} + \dots$$

The first term is just the number of nodes in the molecular graph. The second term is zero because R is traceless. Let us consider the third term then. The diagonal entries of R^2 are $(R^2)_{ii} = \sum_{(i,j) \in E} (k_i k_j)^{-1/2}$, which indicates that

$$(2.9) \quad \text{tr} \frac{(\beta R)^2}{2!} = 2\beta^2 \chi(G).$$

Then, when $\beta \rightarrow \epsilon$ only the first few terms of the series expansion survive, which proves the result. \square

Remark 3. Notice that the Randić index can be obtained as the sum of the squares of the eigenvalues of R , i.e., $\chi(G) = \frac{1}{2} \sum_{j=1}^n \sigma_j^2$.

3. WHY THE RANDIĆ INDEX WORKS IN PREDICTING PHYSICAL PROPERTIES?

We have found in the previous section that the Randić index is directly related to the partition function of a TBH when T is sufficiently high. A connection between the partition function and usual thermodynamic quantities can be obtained through the free energy $F = -\beta^{-1} \ln Z$. This is obtained through the chemical potential μ [32–34]:

$$(3.1) \quad \mu = -\beta^{-1} \left(\frac{\partial \ln Z}{\partial N} \right)_{T,V},$$

where N is the number of molecules. Let us consider that μ^{liq} and μ^{vap} are the chemical potentials of the substance in the liquid and vapors states, respectively. Let us consider that the molecules are partitioned between the liquid and vapor phases at the same temperature and pressure, i.e., there is an equilibrium between the two phases. Then,

$$(3.2) \quad \mu^{\text{liq}} = \mu^{\text{vap}},$$

which implies that, because N , T and V are the same, $Z^{\text{liq}} = Z^{\text{vap}}$. This equilibrium must occur at a sufficiently high temperature. Then, let $0 < \epsilon \ll 1$ be a value of β . Suppose that the equilibrium between the liquid and vapor phases occurs when $\beta \rightarrow \epsilon$. Let us also consider that the vapor phase can be approximated as an ideal gas. Then it is known that the chemical potential is

$$(3.3) \quad \mu^{\text{vap}} = \beta^{-1} \ln(\beta P Z^{\text{vap}}),$$

where P is the pressure of the gas. We can write this expression as

$$(3.4) \quad \mu^{\text{vap}} = \beta^{-1} \ln(\beta P) + \beta^{-1} \ln(Z^{\text{vap}}),$$

Therefore, when $\beta \rightarrow \epsilon$

$$(3.5) \quad \mu_{\beta \rightarrow \epsilon}^{\text{vap}} = \mu_{\beta \rightarrow \epsilon}^{\text{liq}} = \beta^{-1} \ln(\beta P) + \beta^{-1} \ln(n + \beta^2 \chi),$$

where the last term is the free energy of the system: $F_{\beta \rightarrow \epsilon} = -\beta^{-1} \ln(n + \beta^2 \chi)$.

The previous relations indicate that at sufficiently high temperature, the properties of a given substance can be directly related to the Randić index via the chemical potential. Therefore, this clearly explains on physico-chemical grounds why this index has been so successful in modeling and predicting certain physical properties of organic molecules.

3.1. Microscopic analysis. We can also figure out what happens at the microscopic level on the molecular graph using the current formalism. At every node i of the molecular graph there are k_i edges pointing out from the node with weights corresponding to the inverse square root of the degree of i . Also there are k_i edges pointing to that node, with weights equal to $k_j^{-1/2}$, such that $(i, j) \in E$. Therefore, the out-degree of i is $k_i^{out} = k_i^{-1/2}$ and its in-degree is $k_i^{in} = \sum_{(i,j) \in E} k_j^{-1/2}$. Consequently, in non-regular graphs there are nodes with a disbalance between the in- and out-degree as measured by $\Delta k_i = k_i^{in} - k_i^{out}$. That is, we consider the contribution of out-degree as negative (particles leaving that node) and the in-degree as positive (particles arriving to that node). Let us consider the case of the two isomers of butane. In the case of n-butane the methyl groups have $\Delta k_{CH_3} = (1 - \sqrt{2})/\sqrt{2}$ while the methylene groups have: $\Delta k_{CH_2} = (\sqrt{2} - 1)/\sqrt{2}$. That is, while the methyl groups have negative values, the methylene have positive ones. This coincides with the known fact that methyl groups are electron-donors due to their inductive effects. That is, the methyl groups “push” the electrons by their inductive effects towards the methylene ones. Therefore, in the case of iso-butane the values are $\Delta k_{CH_3} = (1 - \sqrt{3})/\sqrt{3}$ and $\Delta k_{CH} = (\sqrt{3} - 1)/\sqrt{3}$, displaying more inductive effects from the methyl groups than in n-butane. Such bigger imbalance in iso-butane, where electrons are more heavily pushed from the methyl groups may produce Coulombic interatomic repulsions between molecules in the liquid state which drop their boiling point as observed experimentally and predicted by the Randić index. This is exactly what it is observed when we study the linear correlations between $Z_R(\beta)$ and the boiling points ($^{\circ}C$) of the 18 acyclic isomers of octane. As can be seen in Fig. 3.1 (a) the best correlation for the boiling point is obtained for $Z_R(\beta)$ when $\beta \rightarrow \epsilon$, which corresponds to a very high temperature regime.

3.2. How to improve a QSPR?. It is then interesting to ask what should be the best correlation between $Z_R(\beta)$ and a property of octane isomers which characterizes their properties at $\beta \gg 0$. That is, while the boiling temperature, which occurs at high temperature, is well described by $Z_R(\beta \rightarrow \epsilon)$, what it would be expected for a property, such as the molar volume in cm^3/mol at $20^{\circ}C$. Molar volume (MV) is the ratio of the molar mass and the density of the substance. Because the density is dependent on the temperature we should obtain MV at a given T . The linear correlation coefficient of the correlation between the Randić index and MV is $r \approx 0.013$. That is, the Randić index is very bad at describing the molar volume of octane isomers. On the light of the current approach we can say that MV is not well-described by $\chi(G)$ because MV is a property measured at $\beta \gg 0$, while the Randić index is describing processes at $\beta \rightarrow \epsilon$. Therefore, we can investigate what happen if we instead consider $Z_R(\beta)$ for $0 < \beta < \infty$. The results are illustrated in Fig. 3.1 (b) where it can be seen that the Pearson correlation coefficient for this property dramatically increases up to 0.964 for $\beta \approx 2.4$. Obviously, when $\beta \rightarrow \epsilon$ we approach the very bad correlation obtained with the “standard” Randić index. This example illustrates how the current approach not only allows to interpret physically this topological index but also allow to improve significantly some of the QSPR models obtained with it.

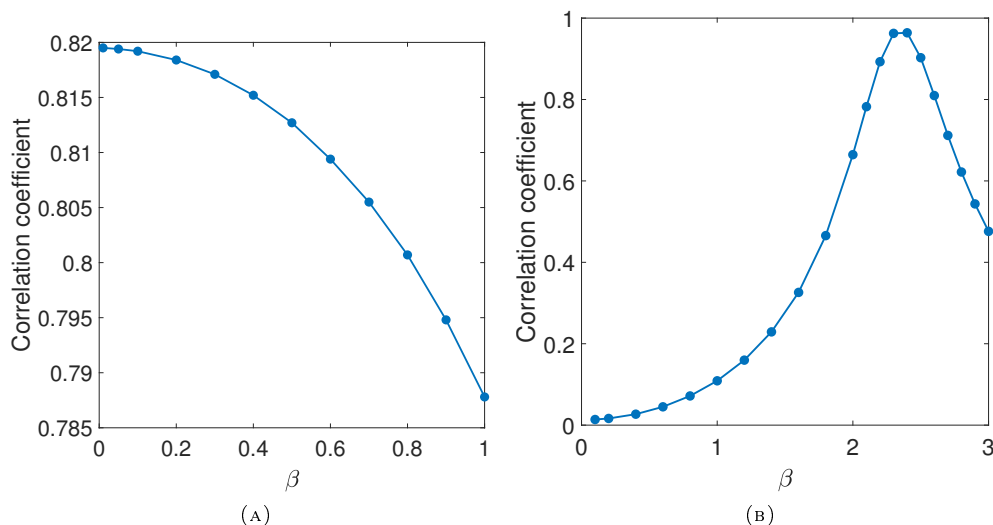


FIGURE 3.1. Pearson correlation coefficient of the linear correlation between $Z_R(\beta)$ vs. the boiling point ($^{\circ}\text{C}$) (a) and the molar volume (cm^3/mol) at 20°C (b) of all acyclic octane isomers for different values of the inverse temperature β .

4. OTHER TOPOLOGICAL INDICES

There are other topological indices which can also be derived analytically from the current theoretical framework. The first of them is the so-called second Zagreb index M_2 [35, 36]. In this case, let us define the following Hamiltonian

$$(4.1) \quad \mathcal{H}_M := - \sum_{ij} k_j c_i^{\dagger} c_j A_{ij},$$

where we have written k_j instead of k_i to make that higher degree vertices pull the electrons toward them. That is, in a bond with atoms having degrees k_i and k_j , such that $k_j > k_i$ the electrons are pulled by atom j with “force” equal to k_j , while they are pulled toward i with force k_i . Hence,

$$(4.2) \quad \mathcal{H}_M := -AK,$$

and $M_2 = \sum_{(i,j) \in E} (k_i k_j) = \lim_{\beta \rightarrow \epsilon} \frac{1}{\beta^2} (Z_M - n)$, where $Z_M = \text{tr}(e^{\beta AK})$.

Another popular index is the so-called Balaban J index [37, 38], which is very similar to the Randić index in the shape of its invariant. Let us define $s_i = \left(D \mathbf{1}^T \right)_i$ where D is the shortest-path (topological) distance matrix of G . Then, let us define the following Hamiltonian

$$(4.3) \quad \mathcal{H}_B := - \sum_{ij} s_i^{-1/2} c_i^{\dagger} c_j A_{ij},$$

where the atom i pulls the electrons toward it with a force equal to $s_i^{-1/2}$. That is, an atom which has relatively small shortest path distance with the rest of atoms in the molecule attracts more the electrons toward it than an atom which is far from the rest. Therefore, we can write

$$(4.4) \quad \mathcal{H}_B := -S^{-1/2}A,$$

where $S = \text{diag}(D\mathbf{1}^T)$. Then, $J = \sum_{(i,j) \in E} (s_i s_j)^{-1/2} = \lim_{\beta \rightarrow \epsilon} \frac{1}{\beta^2} (Z_B - n)$, where $Z_B = \text{tr}(e^{-\beta \mathcal{H}_B})$.

Let us now consider an index known as the “atom-bond-connectivity” or ABC index [39–41], which is defined as

$$(4.5) \quad ABC = \sum_{(v,w) \in E} \left(\frac{k_i + k_j - 2}{k_i k_j} \right)^{1/2}.$$

We can consider that the parameter t_{ij} used in the case of the Randić index is written as $(1/k_i^2)^{1/4}$. Then, we can write the following modified TBH:

$$(4.6) \quad \mathcal{H}_{ABC} := - \sum_{ij} \left(\frac{\delta(e_{ij})}{k_i^2} \right)^{1/4} c_i^\dagger c_j A_{ij},$$

where $\delta(e_{ij}) = k_i + k_j - 2$ is the degree of the bond $(i, j) \in E$. In order to understand the main difference between the hopping parameter used for ABC and the one used in the Randić index let us consider the molecular graph of 2-methylbutane as illustrated in Fig. 4.1. The TBH of the Randić index considers that the CH group pushes the electrons equally towards the methyl groups as to the methylene one, i.e., the weights of the three bonds departing from the CH group has the same weigh of $3^{-1/2}$. Similarly, the methylene group (CH_2) pushes the electrons toward the CH and the methyl group with equal force, i.e., $2^{-1/2}$. In the ABC index, however, the CH group pushes the electrons more strongly towards the methylene group—which is more electronegative—than towards the methyl groups. In a similar fashion, the methylene group pushes the electrons more strongly towards the CH group than towards the methyl one. This better account of the electronegativity in carbon atoms of alkanes could be responsible for the nice properties of this index in predicting the heat of formation of these compounds [44].

Let us now introduce the edge-degree weighted adjacency matrix $\tilde{A}(G)$ as the square symmetric matrix whose entries are defined as

$$(4.7) \quad \tilde{A}_{ij} = \begin{cases} (k_i + k_j - 2)^{1/4} & \text{if } (i, j) \in E \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$(4.8) \quad \mathcal{H}_{ABC} := -K^{-1/2} \tilde{A}.$$

Therefore, $ABC = \lim_{\beta \rightarrow \epsilon} \frac{1}{\beta^2} (Z_{ABC} - n)$, where $Z_{ABC} = \text{tr}(e^{-\beta K^{-1/2} \tilde{A}})$.

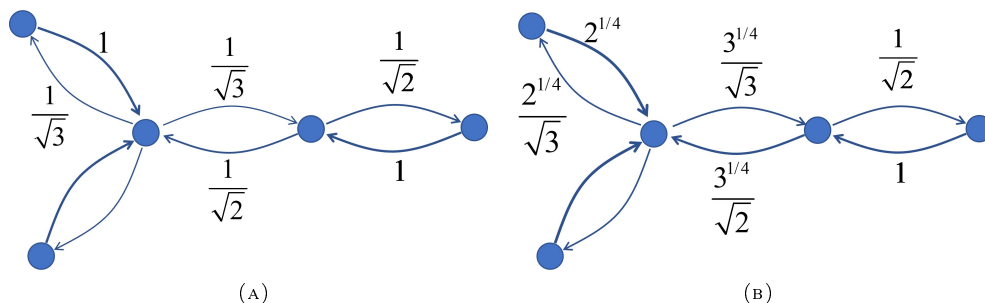


FIGURE 4.1. Illustration of the different weighting scheme given to the bonds in 2-methylbutane by the tight-binding Hamiltonians used here for the Randić (a) and the ABC (b) indices.

Remark 4. The matrices AK , $S^{-1/2}A$, $K^{-1/2}\tilde{A}$, previously defined, are non-Hermitian diagonalizable ones, as can be proved in an analogous way as done in Lemma 1.

Remark 5. The so-called first Zagreb index $M_1 = \sum_{i \in V} k_i^2$ [35, 36] can be written as $M_1 = \sum_{(ij) \in E} (k_i + k_j)$, such that we can write a tight-binding Hamiltonian that also describes this index:

$$(4.9) \quad \mathcal{H}_{M_1} := - \sum_{ij} \sqrt{k_i + k_j} c_i^\dagger c_j A_{ij}.$$

Similarly, the Hamiltonian

$$(4.10) \quad \mathcal{H}_W := - \sum_{ij} \sqrt{\frac{s_i}{k_i} + \frac{s_j}{k_j}} c_i^\dagger c_j A_{ij},$$

will account for the Wiener index [42, 43]: $W := \frac{1}{2} \sum_{i \in V} s_i$.

5. OTHER STATISTICAL-MECHANICAL PARAMETERS

We have previously defined the probability P_j that the molecular graph is found in a state with energy E_j (see eq. (2.5)). We can use this definition for all the Hamiltonians defined in this work. That is, let $\mathcal{H}_{\text{type}}$ be any of the Hamiltonians defined here, i.e., $\text{type} = \{R, M, B, ABC\}$. Then, if E_j^{type} are the corresponding energies defined by such Hamiltonians, i.e., the negative of the eigenvalues σ_j^{type} of the corresponding matrices, we have

$$(5.1) \quad P_j^{\text{type}} = \frac{e^{-\beta E_j^{\text{type}}}}{\sum_j e^{-\beta E_j^{\text{type}}}} = \frac{e^{\beta \sigma_j^{\text{type}}}}{Z^{\text{type}}},$$

as the probabilities of finding the molecular graph in the corresponding j th state defined by $\mathcal{H}_{\text{type}}$. The terms $Z^{\text{type}} = \sum_j e^{\beta \sigma_j^{\text{type}}}$ are new examples of the so-called

Estrada index (the reader is referred to the recent review [45] for an account of the literature).

Then, using Shannon definition of entropy (see [46] for a similar approach in the context of complex networks) we can find:

$$(5.2) \quad S^{type} := -k_B \sum_j P_j^{type} \ln P_j^{type},$$

which can be written as

$$(5.3) \quad \begin{aligned} S^{type} &= -k_B \sum_j P_j^{type} [\beta \sigma_j^{type} - \ln Z^{type}] \\ &= -\frac{1}{T} \sum_j P_j^{type} \sigma_j^{type} + k_B \ln Z^{type}. \end{aligned}$$

Then, multiplying both side of the last equation by T we get

$$(5.4) \quad -\beta^{-1} \ln Z^{type} = -\sum_j P_j^{type} \sigma_j^{type} - TS^{type},$$

which by using the well-known expression $F = H - TS$, clearly identifies the enthalpy H^{type} and free energy F^{type} of the molecular graph in the context of the TBH used here:

$$(5.5) \quad \begin{aligned} H^{type} &= -\sum_j P_j^{type} \sigma_j^{type}, \\ F^{type} &= -\beta^{-1} \ln Z^{type}. \end{aligned}$$

Therefore, when $\beta \rightarrow \epsilon$, such that $Z^{type} \approx n + \epsilon^2 \mathcal{I}^{type}$, where \mathcal{I}^{type} is the corresponding TI, we have

$$(5.6) \quad \begin{aligned} H^{type} &\approx -\sum_j \frac{\sigma_j^{type} e^{\epsilon \sigma_j^{type}}}{n + \epsilon^2 \mathcal{I}^{type}} \\ F^{type} &\approx -\beta^{-1} \ln (n + \epsilon^2 \mathcal{I}^{type}), \end{aligned}$$

from which also the entropy can be obtained using: $S^{type} = \frac{1}{T} (H^{type} - F^{type})$. In this way the TI are directly related to statistical mechanical properties of the molecular graphs.

6. CONCLUSIONS

The main conclusion of this work is that TI are the partition functions of molecules on specific tight-binding approaches. In particular, they are the partition functions of molecules submerged into thermal baths at very high temperature when the molecular electronic properties are described by specific TBH. The specificity of such TBH is that their hopping parameters are given by topological properties of the atom-atom interactions, such as the degrees of the nearest neighbors or related properties based on topological distances. Here we have derived such first-principles approach to the Randić, Zagreb, Balaban, Wiener and ABC indices, but it can also

be extended to others indices using the same approach. The current work opens new avenues for the study of topological invariants in chemistry. We can mention for instance the study of statistical mechanical descriptors derived from topological TBH, the application of the new partition functions as molecular descriptors in QSPR using different values of β , the analysis of algebraic properties of the new matrices representing the topological TBH, among others.

ACKNOWLEDGMENT

The author thanks financial support from the grant PID2019-107603GB-I00 of the MCIN/AEI /10.13039/501100011033/.

REFERENCES

- [1] Strogatz, S. H. Exploring complex networks. *Nature* 410 (2001) 268-276.
- [2] Estrada, E. The structure of complex networks: theory and applications. Oxford University Press, 2012.
- [3] Latora, V., Nicosia, V. and Russo, G. Complex networks: principles, methods and applications. Cambridge University Press, 2017.
- [4] Rouvray, D. H. The pioneering contributions of Cayley and Sylvester to the mathematical description of chemical structure. *J. Mol. Struct.: THEOCHEM* 185 (1989) 1-14.
- [5] Trinajstić, N. Chemical Graph Theory. CRC press, 2018.
- [6] Wagner, S., and Wang, H. Introduction to Chemical Graph Theory. Chapman and Hall/CRC, 2018.
- [7] Devillers, J. and Balaban, A. T. (Eds.) Topological Indices and Related Descriptors in QSAR and QSPR. Gordon and Breach Science Publishers: Singapore. 1999.
- [8] Estrada, E., and Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* 8 (2001) 1573-1588.
- [9] Li, X., and Shi, Y. A survey on the Randić index. *MATCH Commun. Math. Comput. Chem* 59 (2008) 127-156.
- [10] Randić, M. Characterization of molecular branching. *J. Am. Chem. Soc.* 97 (1975) 6609-6615.
- [11] Randić, M. The connectivity index 25 years after. *J. Mol. Graph. and Modell.* 20 (2001) 19-35.
- [12] Dearden, J. C. The use of topological indices in QSAR and QSPR modeling. *Advances in QSAR Modeling*. Springer, Cham, 2017, pp. 57-88.
- [13] Estrada, E., and Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* 8 (2001) 1573-1588.
- [14] Gonzalez-Diaz, H., Vilar, S., Santana, L., & Uriarte, E. Medicinal chemistry and bioinformatics-current trends in drugs discovery with networks topological indices. *Curr. Top. Med. Chem.*, 7 (2007) 1015-1029.
- [15] Dudek, A. Z., Arodz, T. and Gálvez, J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb. Chem. & High Throug. Screen.* 9 (2006) 213-228.
- [16] Galvez, J., Garcia-Domenech, R., de Julián-Ortiz, J. V., & Soler, R. Topological approach to drug design. *J. Chem. Inf. Comput. Sci.*, 35 (1995) 272-284.
- [17] Todeschini, R., and Consonni, V. Handbook of Molecular Descriptors. John Wiley & Sons, 2008.

- [18] Guimera, R., et al. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Nat. Acad. Sci.* 102 (2005) 7794-7799.
- [19] Sporns, O. *Networks of the Brain*. MIT press, 2010.
- [20] Borgatti, S. P., Everett, M. G. and Johnson, J. C. *Analyzing Social Networks*. Sage, 2018.
- [21] Randić, M. On history of the Randić index and emerging hostility toward chemical graph theory. *MATCH Commun. Math. Comput. Chem.* 59 (2008) 5-124.
- [22] Todeschini, R., Cazar, R. and Collina, E. The chemical meaning of topological indices. *Chemom. Intell. Lab. Syst.* 15 (1992) 51-59.
- [23] Estrada, E. Generalization of topological indices. *Chem. Phys. Lett.* 336 (2001) 248-252.
- [24] Basak, S. C., et al. Topological indices: their nature, mutual relatedness, and applications. *Math. Modell.* 8 (1987) 300-305.
- [25] Randić, M., and Zupan, J. On interpretation of well-known topological indices. *J. Chem. Infor. Comp. Sci.* 41 (2001) 550-560.
- [26] Labanowski, J. K., Motoc, I. and Dammkoehler, R. A. The physical meaning of topological indices. *Comp. Chem.* 15 (1991) 47-53.
- [27] Goringe, C. M., Bowler, D. R. and Hernandez, E. Tight-binding modelling of materials. *Rep. Progr. Phys.* 60 (1997): 1447-1512.
- [28] Yang, W., and Mak, C. H. Transferable tight-binding potential for hydrocarbons. *Chem. Phys. Lett.* 235 (1995) 37-46.
- [29] Pan, B. C. Tight-binding potential for hydrocarbons. *Phys. Rev. B* 64 (2001) 155408.
- [30] Horsfield, A. P., et al. Computational materials synthesis. I. A tight-binding scheme for hydrocarbons. *Phys. Rev. B* 54 (1996) 15773.
- [31] Andrews, F. C. *Equilibrium Statistical Mechanics*. John Wiley & Sons., New York, 1963.
- [32] Buchwald, P., and Bodor, N. Molecular size-based model to describe simple organic liquids. *J. Phys. Chem. B* 102 (1998) 5715-5726.
- [33] Mutelet, F., and Rogalski, M. Estimating thermodynamic properties of organic compounds using a solvation model. *Phys. Chem. Chem. Phys.* 3 (2001) 432-436.
- [34] Ben-Naim, A. Y. *Solvation Thermodynamics*. Springer Science & Business Media, 2013.
- [35] Zhou, B. Zagreb indices. *MATCH Commun. Math. Comput. Chem* 52 (2004) 13-118.
- [36] Gutman, I., and Trinajstić, N. Graph theory and molecular orbitals. Total π -electron energy of alternant hydrocarbons. *Chem. Phys. Lett.* 17 (1972) 535-538.
- [37] Balaban, A. T. Highly discriminating distance based numerical descriptor. *Chem. Phys. Lett* 89 (1982) 399-404.
- [38] Balaban, A. T. Topological indices based on topological distances in molecular graphs. *Pure Appl. Chem.* 55 (1983) 199-206.
- [39] Estrada, E., Torres, L., Rodriguez, L., & Gutman, I. An atom-bond connectivity index: modelling the enthalpy of formation of alkanes. *Indian J. Chem.* 37A (1998) 849-855.

- [40] Estrada, E. Atom–bond connectivity and the energetic of branched alkanes. *Chem. Phys. Lett.* 463 (2008) 422-425.
- [41] Estrada, E.. The ABC matrix. *J. Math. Chem.* 55 (2017) 1021-1033.
- [42] Nikolić, S., and Trinajstić, N. The Wiener index: Development and applications. *Croat. Chem. Acta* 68 (1995) 105-129.
- [43] Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* 69 (1947) 17-20.
- [44] Gutman, I., Tošović, J., Radenković, S., & Marković, S. On atom-bond connectivity index and its chemical applicability. *Indian J. Chem.* 51A (2012) 690-694.
- [45] Estrada, E. The many facets of the Estrada indices of graphs and networks. *SeMA J.* (2021) 1-69.
- [46] Estrada, E., and Hatano, N. Statistical-mechanical approach to subgraph centrality in complex networks. *Chem. Phys. Lett.* 439 (2007) 247-251.

INSTITUTE FOR CROSS-DISCIPLINARY PHYSICS AND COMPLEX SYSTEMS (IFISC, UIB-CSIC),
CAMPUS UNIVERSITAT DE LES ILLES BALEARS E-07122, PALMA DE MALLORCA, SPAIN. **E-MAIL:**
ESTRADA@IFISC.UIB-CSIC.ES