Current Bioinformatics. 2015; 10 (5): 692-713

# MI-NODES multiscale models of metabolic reactions, brain connectome, ecological, epidemic, world trade, and legal-social networks

Aliuska Duardo-Sanchez[1,2], Humberto González-Díaz[3,4] and Alejandro Pazos[l]

[1] *Department of Information and Communication Technologies, University of A Coruña UDC, 15071, A Coruña, Spain*
[2] *Department of Especial Public Law, Financial and Tributary Law Area, Faculty of Law, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain*
[3] *Department of Organic Chemistry II, University of the Basque Country UPV/EHU, 48940, Leioa, Spain*

**Abstract**
Complex systems and networks appear in almost all areas of reality. We find then from proteins residue networks to Protein Interaction Networks (PINs). Chemical reactions form Metabolic Reactions Networks (MRNs) in living beings or Atmospheric reaction networks in planets and moons. Network of neurons appear in the worm C. elegans, in Human brain connectome, or in Artificial Neural Networks (ANNs). Infection spreading networks exist for contagious outbreaks networks in humans and in malware epidemiology for infection with viral software in internet or wireless networks. Social-legal networks with different rules evolved from swarm intelligence, to hunter-gathered societies, or citation networks of U.S. Supreme Court. In all these cases, we can see the same question. Can we predict the links based on structural information? We propose to solve the problem using Quantitative Structure-Property Relationship (QSPR) techniques commonly used in chemo-informatics. In so doing, we need software able to transform all types of networks/graphs like drug structure, drug-target interactions, protein structure, protein interactions, metabolic reactions, brain connectome, or social networks into numerical parameters. Consequently, we need to process in alignment-free mode multitarget, multiscale, and multiplexing, information. Later, we have to seek the QSPR model with Machine Learning techniques. MI-NODES is this type of software. Here we review the evolution of the software from chemoinformatics to bioinformatics and systems biology. This is an effort to develop a universal tool to study structure-property relationships in complex systems.

**Keywords:**
QSPR models in complex networks; Drug-target networks, Metabolic networks, Brain connectome, Social networks, World trade, US supreme court citation networks, Spain's financial law.

## 1. INTRODUCTION

### 1.1. Structure-Property Problem in Complex Systems

Complex systems and networks appear in phenomena belonging to almost all areas of reality at very different temporal and spatial scales [1]. We find them from bio-molecular structures of proteins residue networks [2] to Protein Interaction Networks (PINs) [3]. The coupling of chemical reactions lead to the formation of Metabolic Reactions Networks (MRNs) [4, 5] in living beings or Atmospheric Reaction Networks (ARNs) in planets and moons like Earth, Mars, Venus, and Titan [6]. Complex patterns appear in the network of neurons of the worm C. *elegans* [7], in Human Brain Connectome [8], or in Artificial Neural Networks (ANNs) [9]. Spreading patterns appears for contagious outbreaks networks in humans [10] or in malware epidemiology due to infection with viral software in internet or Wi-Fi wireless networks [11, 12]. Complex behavior emerge from basic rules of Swarm Intelligence (SI) [13], collaboration in hunter-gathered societies [14], or in legislation code in the citation network of U.S. Supreme Court (USSC), as well [15]. The most basic issues are structural: how does one characterize the connectivity patterns in those networks? Are there any unifying features underlying their topology? Different research groups have begun to shed light over these unifying aspects of the structure and dynamics of complex networks indeed [1, 7, 16-21]. Networks are represented by means of a graph as a way to capture essential information. Graphs in turn are sets of items, drawn as dots, or *nodes*, interconnected by lines or arcs, which represents wires, ties, links, edges, bonds, *etc.* Consequently, the nodes can represent atoms, molecules, proteins, nucleic acids, drugs, cells, organisms, parasites, people, words, laws, computers or any other part of a real system. Moreover, the links represent relationships between the nodes such as chemical bonds, physical interactions, metabolic pathways, pharmacological action, law recurrence or social ties [4, 22-30].

Tenazinha and Vinga [31] reviewed frameworks currently available for modeling and analyzing integrated biological networks, in particular metabolic, gene regulatory and signaling networks. In effect, there are different experimental and/or theoretical methods to assign no de-no de links depending on the type of network we want to construct. Unfortunately, many of these methods are expensive in terms of time or resources (especially the experimental ones). In addition, different methods to link nodes in the same type of network are not very accurate in such a way that they do not always coincide. One possible solution to this problem is the use of Quantitative Structure-Property (QSPR) models. Traditionally, this methodology has been used in chemo-informatics. Most often, QSPR-like models use as input structural parameters derived from the graph representation of the network-like system under study [32]. Many authors refer to the numerical parameters of a graph as Topological Indices (TIs); mainly in the case of global studies (properties of full system). We can use local TIs of a sub-graph or centralities $C_t(j)$ of type t for the $j^{th}$ node in the network to study a local property of a given part of the system [33-35]. In Table 1, we present the names, symbol, formula, and software used to ca1culate some of these centralities.

In order to develop such computational models we need to use modeling techniques to process chemical information from public databases. These databases have accumulated immense datasets of experimental results of pharmacological trials for many compounds. For instance, STITCH [36-38], TTD [39], Super Target [40, 41], or the colossal CHEMBL lists thousands of drugs, targets, and drug-target interactions. This huge amount of information offers a fertile field for the application of computational techniques [42, 43]. The analysis of all this data is very complex due to different features of the chemical and pharmacological information present: (1) multi-scaling, (2) multi-targeting, (3) alignment dependent, and/or (4) multi-output or plexing nature. The same features appear in biological, technological, social, and other complex networks.

**Table 1**. MI-NODES vs some classic node centralities

| Name | Formula | Software | Ref. |
|---|---|---|---|
| | | | |
| Degree | $C_{\deg}\ (j) = \deg(j)$ | CBI | [34] |
| Eccentricity | $C_{ecc}(j) = \max\{dist(i,j)\}^{-1}$ | CBI | |
| Closeness | $C_{clo}(j) = \left(\sum_{j\epsilon V} dist(i,j)\right)^{-1}$ | CBI | |
| Radiality | $C_{rad}(j) = \sum_{w\epsilon V} (\Delta_G + 1 - dist(i,j))\,/(n-1)$ | CBI | |
| Centroid Values | $C_{cen}(j) = \min\{f(i,j): i\epsilon V\{j\}\}$ | CBI | |
| Stress | $C_{str}(j) = \sum_{s\notin v\epsilon V} \sum_{t\notin v\epsilon V} \sigma_{st}(j)$ | CBI | |
| Shortest-path Betweenness | $C_{spb}(j) = \sum_{s\notin v\epsilon V} \sum_{t\notin v\epsilon V} \delta_{st}(j)$ | CBI | |
| Current-Flow Closeness | $C_{cfc}(j) = (n-1)/\left(\sum_{i\notin V} p_{ji}(j) - p_{ij}(i)\right)$ | CBI | |
| Current-Flow Betweenness | $c_{cfb}(j) = \sum_{s,t\epsilon V} \tau_{st}(j)/(n-1)(n-2)$ | CBI | |
| Katz Status Index | $C_{katz} = \sum_{k=1}^{\infty} \alpha^k \cdot (A^t)^k \cdot u$ | CBI | |
| Eigenvector | $EC(j) = e_1(j)$ | CBI | |
| Closeness Vitality | $C_{clv}(j) = W(G) - W(G\{j\})$ | CBI | |
| Markov-Randic | ${}^kC_\chi(j) = \sum_{i}^{\delta_j} (\delta_i \cdot \delta_j)^{1/2} \cdot {}^kp_{ij}$ | MI | [46] |
| Markov-Shannon entropy | ${}^kC_\theta(j) = -\sum_{i}^{n} ({}^kp_j) \cdot \log({}^kp_j)$ | MI | [47] |
| Markov Spectral moments | ${}^kC_\pi(j) = \sum_{i=j}^{n} {}^kp_{ij} = [l(\sqcap)^k]$ | MI | [48 |
| Markov-Harary | ${}^kC_H(j) = \frac{1}{2}\sum_{i}^{\delta_j} {}^kp_{ij}^{-1}$ | MI | [49] |
| Markov-Galvez | ${}^kC_G(j) = \frac{1}{2}\sum_{i,j}^{n} \left|{}^kCT_{ij}\right| \cdot \delta_j$ | MI | [50] |
| Markov-Rucker | ${}^kC_{WC}(j) = \frac{1}{2}\sum_{i}^{\delta_j} {}^kp_{ij}$ | MI | [51] |
| Markov-BM Autocorrelation | ${}^kC_{BM}(j) = \frac{1}{2}\cdot\sum_{i}^{\delta_j} {}^kp_{ij} \cdot {}^kp_{ij}$ | MI | [52] |
| Markov-Wiener | ${}^kC_w(j) = \frac{1}{2}\cdot\sum_{i\to l}^{\delta_j} {}^kp_{ij} \cdot d_{ij}$ | MI | [53] |
| Markov-Balaban | ${}^kC_J(j) = \frac{q}{\mu+1}\cdot\sum_{i\to l}^{\delta j} ({}^kp_{ij} \cdot S_i \cdot S_j)^{-1/2}$ | MI | - |

[a] All symbols used m these formulae are very common m networks literature and cannot be explamed m detall here. However, *G* IS an undirected or directed graph with $n=|V|$ vertices; deg(*v*) denotes the degree of the vertex v in an undirected graph; dist(*v*, *w*) denotes the length of a shortest path between the vertices *v* and *w*; $\sigma_{st}$ denotes the number of shortest paths from *s* to *t* and $\sigma_{st}(v)$ the number of shortest path from s to *t* that use the vertex *v*. O and A are the topological distance and the adjacency matrix of the graph *G*. Please, for more details see the references cited and others.

### *1.2. Why Do We Need Multiscale Models?*

One of the more important characteristics enumerated before is the multi-scale nature of many important problems. Currently, the use of QSPR-like models in which the inputs are graph parameters is not limited to the study of molecules and has been extended to other complex systems [44]. As we mentioned in the previous paragraph, in multi-target modeling we need to incorporate information about the drug and different molecular targets (proteins, RNA, gene). In this case, we can solve the problem using molecular descriptors. However, in the case of not molecular complex networks we are out of the chemical scale. We can find complex systems formed by networks in many different scales. In general, these scales may be c1assified as time and spatial scales. In the case of time scales, we can find different dynamic networks in a same or different problem that change the pattern of links in different time scales (seconds, min, hours, days, years, or seasons). In this case, we can still circumvent the problem with MA models like those of Bob and Jenkins mentioned before [45].

### *1.3. Why Do We Need Alignment-Free Models?*

Alignment-based and alignment-free methods are two fundamentally different methods used to compare sequences, and genomes by extension [54]. This approach is very useful but only when we found a high homology between the query and the template sequences deposited in the data base and therefore may fail in case of low homology [55]. The lack of function annotation (defined biological function) for the best alignment matches is another cause for alignment pitfalls [56]. Yet, functional information - either experimentally validated or computationally inferred by similarity - remains completely missing for approximately 30% of human proteins [57]. In 2012, Wood *el al.* [58] analyzed 1,474 prokaryotic genome annotations in GenBank. They identified 13,602 likely missed genes that are homologues to non- hypothetical proteins. It is very relevant that they also identified 11,792 likely missed genes that are homologues only to hypothetical proteins, despite evidence of their protein-coding nature. Alignment approaches also views proteins and nuc1eic acids as linear sequences of discrete units similar to linguistic representations ignoring 3D structure and overlooks well-documented long-range interactions [59]. On the other hand, alignment-free methods have emerged as a solution to these problems. Vinga and Almeida [60] reviewed two of the more important types of alignment free methods: (1) methods based on word frequency and resolution-free methods. In parallel, Chou [61, 62], Randic [63], González-Díaz [29, 64-70], and others have introduced alignment-free parameters for the pseudo- folding of sequences into geometrically constrained 2D, 3D, or higher dimension spaces using simple heuristics. Pseudo- folding parameters or sequence molecular descriptors codify non-linear relationships without necessity of determination of real 3D structures (graph representations) and are used as inputs of machine learning experiments to seek QSPR models able to predict function from sequence without rely upon alignment [35, 66, 71-74].

### *1.4. Why Do We Need Multi-Target Models?*

Multi-targeting complication emerges due to the existence of multi-target compounds [75-77], which led to the formation of complex networks of drug-target and/or target-target interactions. We can represent target interactions as networks of nodes (proteins, gene, RNAs, miRNAs) interconnected by a link when there is a target-target interaction between two of them. In addition, we can represent drug-target networks as a graph with two type of nodes drugs ($d_i$) and targets ($t_j$) interconnected by links ($L_{ij}$). Barabasi *el al.* [78], constructed a drug-target network based on Food and Drug Administration (FDA) drugs and proteins linked by drug-target binary associations. Yamanishi *el al.* [79] also reported a predictive algorithm to construct drug-target networks. Csermely *el al.* [80] have reviewed the state-of-art and trends on the use of networks, inc1uding drug-target networks, for drug discovery. In general, many of the c1assic models used in chemo-informatics are able to predict the biological activity of some types of drugs against only one target using molecular descriptors of the drug. An alternative is the development of general multi-target models able to predict the interaction ($L_{ij}$) of large libraries of drugs ($d_i$) with a large number of targets (multiple-target models). In this case, we can use molecular descriptors of the drug and the

target. For instance, Viña *et al.* [29] and Prado- Prado *et al.* [81, 82] predicted the different drug-target network using the software MI to ca1culate the structural indices.

### *1.5. Why Do We Need Multiplexing Models?*

However, in multi-plexing modeling we need to use additional operators to incorporate non-structural information. The non-structural information here refers to different assay conditions (c.) like time, concentrations, temperature, cellular targets, tissues, organisms, *etc.* In recent works González-Díaz *et al.*, adapted the idea of Moving Average (MA) operators used in time series analysis with a similar purpose. MA models become popular after the initial works of Bob and Jenkins [45]. In multi-output modeling, we calculate the MA operators as the average of the property of the system (molecular descriptors or others) for all drugs or targets with a specific response in one assay carry out at under a sub-set of conditions (c.), Consequently, our MA operator is not acting over a time domain but over a sub-set of conditions of the pharmacological assays. Botella- Rocamora *el al.* [83], have applied MA of time series theory to the spatial domain, making use of a spatial MA to define dependence on the risk of a disease occurring. The main objective of our work is assessing links in different complex networks. For it, we use MA of properties of nodes of networks (drugs, proteins, reactions, laws, neurons, *etc.)* that form links ($L_{ij}$) in specific sub-set of conditions ($c_j$).

## 2. FROM MARCH-INSIDE TO MI-NODES

In a effort to solve the previous problem, González-Díaz *el al.* introduced the software called MARCH-INSIDE (Markovian Chemicals In Silico Design), or shortly MI, which has become a very useful tool for QSPR studies for drugs, proteins, and complex systems in general [65, 84-97]. MI calculates descriptors ${}^{k}D_{t}(G_{m})$ of type t (entropies, moments, means) and order k for all or some nodes (atoms, aminoacids, nucleic bases) using molecular graph $G_m$ of $m^{th}$ molecule. The graph G represent the ID (sequence), 2D (secondary), or 3D (spatial) structure of a molecular system drug, protein, RNA, artificial polymers, *etc.* [9, 98]. In Fig. (1), we illustrate the user-software interface for classic MI (top) or MI-NODES (bottom).

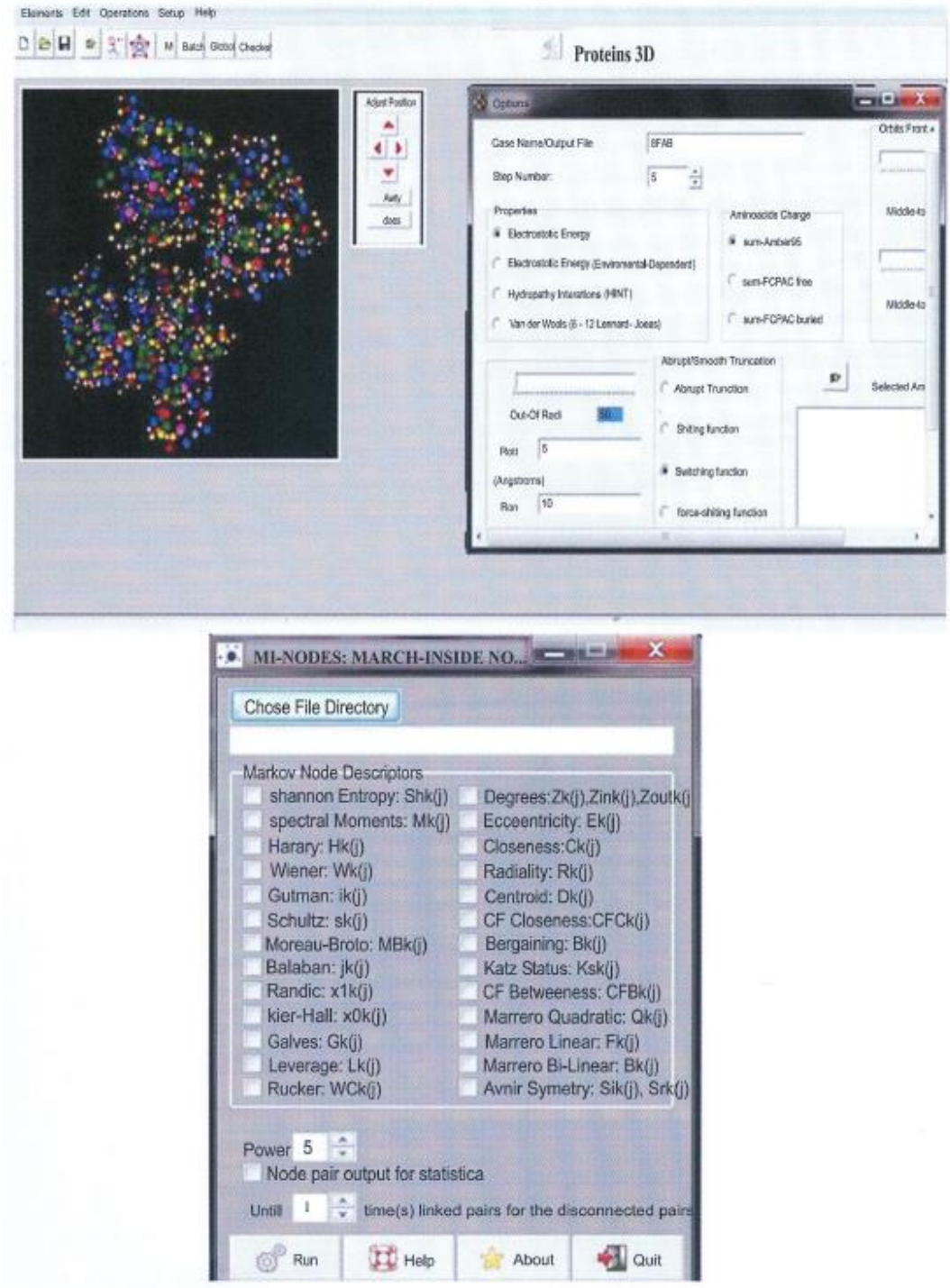



**Fig. (1)**. MI and Mi-NODES user interfaces

However, MI can perform a limited manage of other complex networks. Recently, we have re-programmed the MI application creating a new software application able to manage complex networks. The new program is called MI-NODES (MARCH-INSIDE for NOde DEScriptors) is able to upload files with .mat, .net, and .dat formats and is compatible with other software like Pajek [99] or CentiBin [34]. A very interesting feature of MI-NODES is that it can calculate general versions of classic molecular TIs for large complex networks using Markov Chains theory. In Fig. (2), we show thee general steps used to develop a QSPR model based on the MI algorithm. Briefly, the steps of the MI algorithm are the following.

Step 1 - MI algorithm reads the input files with structural information of the system; essentially nodes, links, and weights;

Step 2 - MI creates a node-node connectivity or adjacency matrix **A**, if not uploaded in the input file. The elements of **A** are $a_{ij} = 1$ if the node $a_i$ is connected to the node $a_j$ and $a_{ij} = 0$ otherwise;

Step 3 - MI transforms **A** into a weighted matrix **W**. The elements of **W** are $w_{ij} = w_j$ if $a_{ij} = 1$ and $w_{ij} = 0$ otherwise. For molecules, the weights are the atomic electronegativity ($\chi_j$), polarizability ($\alpha_j$), amino acid propensities ($\Omega_j$), *etc.* We set constant weights $w_j = 1$ (reduction to adjacency) or equal to no de degree $w_j = \delta_j$ when we do not know the properties of nodes;

Step 4 - MI transforms **W** into a Markov Matrix ${}^{1}\Pi$ and obtain the natural powers of this matrix ${}^{k}\mathbf{\Pi} = ({}^{l}\mathbf{\Pi})^{k}$ . According to Markov Chains theory, the elements of these matrices ${}^{k}p_{ij}$ are probabilities of short/long-range interactions for pairs of nodes place at topological distances $d_{ij} \leq k$;

Step 5 - MI use the values of ${}^{k}\mathbf{\Pi}$ matrices to calculate different molecular descriptors ${}^{k}D_t(G_m)$ for small molecules. The classic MI can be used for small molecules (drugs, metabolites, *etc.)* or biopolymers (proteins, RNAs, DNA). MI-NODES is used to read the files with the structure of complex networks.

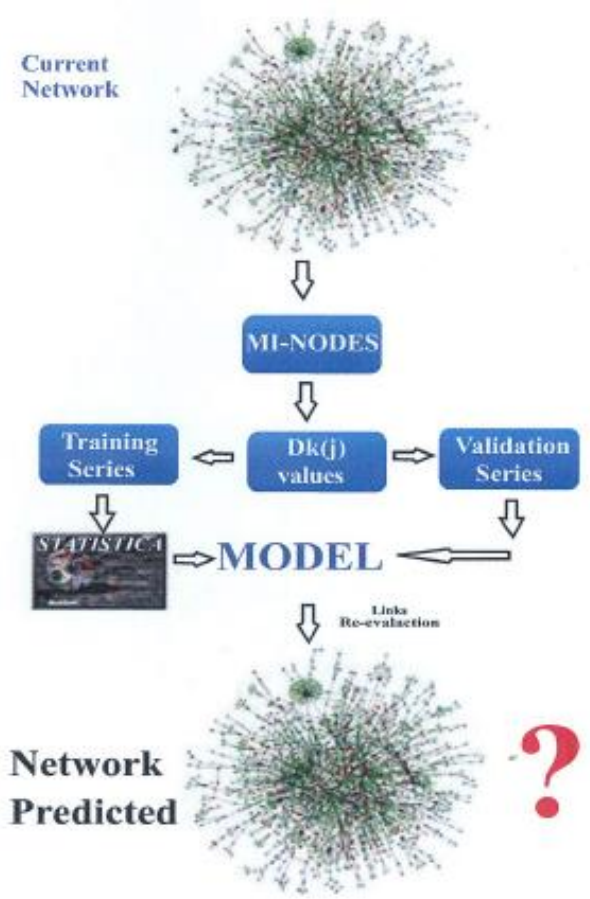



**Fig. (2)**. QSPR analysis of complex networks

## 3. MI PARAMETERS

### *3.1. MI Para meter for Drugs*

MI calculate different types of molecular properties ${}^{k}D_{t}(G_m, w_j)$ [92, 95, 100] based on the molecular graph $G_m$ of the $m^{th}$ molecule and weights of nodes (atoms) equal to physicochemical atomic properties ($w_j$).

We can omit $w_j$ in the notation when we use only one atomic property and declare it *a priori*, *e.g.* $w_j = \chi_j$ the atomic electronegativity. For instance, it is possible to calculate mean atomic electronegativities ${}^{k}D_{\chi}G_m$), Shannon entropy of electron delocalization ${}^{k}D_{\theta}(G_m)$, or spectral moments ${}^{k}D_{\pi}(G_m)$ [91,101].

$$ {}^{k}D_{\chi}(G_m) = \sum_{j \epsilon G} p_k(G_m) \cdot \chi_j \tag{1} $$

$$ {}^{k}D_{\theta}(G_m) = -\sum_{j \epsilon G}^{n} {}^{k}p_j(G_m) \cdot \log\left[ {}^{k}p_j(G_m) \right] \tag{2} $$

$$ {}^{k}D_{\pi}(G_m) = \sum_{i=j \epsilon R}^{n} {}^{k}p_{ij}(G_m) \tag{3} $$

It is possible to consider isolated atoms ($k$ = 0) in a first estimation of the molecular properties ${}^{0}D_{\chi}(G_m)$, ${}^{0}D_{\theta}(G_m)$, or ${}^{0}D_{\pi}(G_m)$. In this case, the probabilities ${}^{0}p_{ij}(w_j)$ are determined without considering the formation of chemical bonds (additive scheme). It is possible to consider the gradual effects of the neighboring atoms placed at distance $k$ using the absolute probabilities $p_k(w_j)$ with which these atoms affect the contribution of the atom $j$ to the molecular property in question.

### *3.2. MI Parameters for Protein 3D Structures*

In the MI algorithm, we codify the information about protein structure using a Markov matrix ${}^{1}\mathbf{\Pi}$ that quantify the probabilities of short-term field interactions among amino acids *(aa)* [9, 72, 102-104]. The matrix ${}^{1}\mathbf{\Pi}$ is constructed as a squared matrix $(n^x n)$, where $n$ is the number of amino *aa* in the m[th] protein with contact map represented by the graph $G_m$ [105-107] In previous works we have predicted protein function based on mean values of 3D-Potentials ${}^{k}D_{\xi}(G_m, E)$, ${}^{k}D_{\xi}(G_m, vdW)$, and ${}^{k}D_{\xi}(G_m, h)$ for different type of interactions or molecular fields derived from ${}^{1}\mathbf{\Pi}$. The main types of the molecular fields used are: Electrostatic (e), van der Waals (vdw), and HINT (h) potentials [106, 108, 109]. The detailed explanation has been published before. In some of these works we calculated also entropy ${}^{k}D_{\xi}(G_m, E)$, ${}^{k}D_{\xi}(G_m, vdW)$, and ${}^{k}D_{\xi}(G_m, h)$ and spectral moment ${}^{k}D_{\theta}(G_m, E)$, ${}^{k}D_{\theta}(G_m, vdW)$, and ${}^{k}D_{\theta}(G_m, h)$ values for the same molecular fields. See the formula of the 3D mean potential, entropy, and moments for the electrostatic field:

$$ {}^{k}D_{\xi}(G_m) = -\sum_{j \epsilon G_i} {}^{k}p_j(G_m) \cdot \xi_0(j) \tag{4} $$

$$ {}^{k}D_{\theta}(G_m) = -\sum_{j \epsilon G}^{n} {}^{k}p_j(G_m) \cdot \log\left[ {}^{k}p_j(G_m) \right] \tag{5} $$

$$ {}^{k}D_{\pi}(G_m) = \sum_{i=j \epsilon G}^{n} {}^{k}p_j(G_m) \tag{6} $$

It is remarkable that the spectral moments depend on the probability ${}^{k}p_{ij}G)$ with which the effect of the interaction f propagates from amino acid i[th] to other neighboring amino acids j[th] and retums to i[th] after k-steps. On the other hand, both the average electrostatic potential and the entropy measures depend on the absolute probabilities ${}^{k}p_j(R)$ with which the amino acid j[th] has an interaction of type f with the rest of *aa*. The software MI [100] performs all these calculations by evaluation of the summation term either for all amino acids or only for some specific groups called regions ($R \epsilon G_m$). We defined the regions in geometric terms and called them as core, inner, middle, or surface region. Please, see details in the literature [9,72,102-104, 109-113].

### *3.3. MI Parameters for Complex Networks*

In previous works, we have introduced new types of MI descriptors ${}^{k}D_t(G_m)$ complex networks. These values can be calculated as the sum of MI no de centralities ${}^{k}C_t(j)$ for each j[th] nodes in the network, see Table 1. These descriptors are Markov chain generalizations of classic TIs. Some of these are Markov-Shannon Entropies [114], Markov-Randić indices [46], or Markov-Harary numbers [115]. We have used Markov- TIs to study several types of complex networks in Biology, Linguistics, Technology, Social, and Legal Sciences. In the next section, we describe different parameters of MI.

We implemented the new centralities in the software MI-NODES (MARCH-INSIDE for NOde DEScriptors) and used it to calculate the node centralities of the networks studied in this work. MI-NODES is a GUI Python/wxPython application developed by our groups. It is an upgrade of part of the code of the software MARCH-INSIDE adapted to manage any kind of complex networks. The program builds a Markov matrix ($^1\mathbf{\Pi}$) for each network using as input the matrix of connectivity or adjacency of nodes often denoted as A. The elements of this stochastic matrix are the node- node transition probabilities *($p_{ij}$)* The probability matrix is raised to the power *k*, resulting $(^1\mathbf{\Pi})^k$ The resulting matrices $^k\mathbf{\Pi}$, which are the $k^{th}$ natural powers of $^1\mathbf{\Pi}$, contain the transition probabilities $^kp_{ij}$. These are the probabilities to reach the $j^{th}$ node moving from the $i^{th}$ node throughout a walk of length k for each *k.* The generalization of the classic TIs and node centralities to general MI indices of order $k^{th}$ is straightforward to realize simply by substitution/multiplication of some parameters used in classic TIs like topological distances ($d_{ij}$) or node degrees (j) by/with Markov matrix parameters like transition probabilities $^kp_{ij}$. We can obtain different MI generalizations of classic TIs and/or no de centralities. For instance, we can calculate k values of the new Markov-Rücker indices $WC_k(G)$ for a graph G (or probabilistic walk counts). We only have to change $d_{ij}$ by $^kp_{ij}$. Conversely, we can obtain k values of new Markov-Wiener indices $W_k(G)$ for a graph G multiplying $d_{ij}$ by $^kp_{ij}$. In so doing, it is possible to run the sum over all nodes in G to calculate global TIs or only over all the r nodes linked to one specific no de *i.* The number of these nodes linked directly to one specific node is equal to $\delta_i$ (the degree of i) and we symbolized here a direct link as $j \rightarrow i$. In a very simple example, we can obtain a total of k values of new Markov-Rücker or probabilistic walk count centralities $WC_k(i)$ for the no de $i^{th}$.

$$^kD_{wc}(G) = \frac{1}{2} \cdot \sum_{i=1}^{D} \sum_{j=1}^{D} {}^kp_{ij}$$

or (7)

$$^kD_{w}(G) = \frac{1}{2} \cdot \sum_{i=1}^{D} \sum_{j=1}^{D} {}^kp_{ij} \cdot d_{ij}$$

$$^kC_{wc}(j) = \frac{1}{2} \cdot \sum_{i=1}^{1} \sum_{j \rightarrow i}^{\delta_i} {}^kp_{ij} = \frac{1}{2} \cdot \sum_{j \rightarrow i}^{\delta_i} {}^kp_{ij}$$

or (8)

$$^kC_{w}(j) = \frac{1}{2} \cdot \sum_{i=1}^{1} \sum_{j \rightarrow i}^{\delta_i} {}^kp_{ij} \cdot d_{ij} = \frac{1}{2} \cdot \sum_{j \rightarrow i}^{\delta_i} {}^kp_{ij} \cdot d_{ij}$$

In Table 1 we list the names, formula, software used for ca1culation, and references of many classic and MI centralities [33, 34, 46, 47, 49-52, 116].

## 4. GENERAL MI MODELS

### *4.1. Models ofDrug-Target Networks (DT-Nets)*

In MI strategy we can use as inputs the parameters of the $m^{th}$ drug molecule or protein ligands with molecular graph ($G_m = L_r$), We use $w_j = \chi_j$ by default, omit it in notations, obtaining the molecular descriptors $^kD_\chi(L_r)$, $^kD_\theta(L_r)$, or $^kD_\pi(L_r)$; by one hand. In addition, we should use the MI parameters of the $s^{th}$ protein sequence or 3D structure to obtain the descriptors $^kD_\chi(P_s)$, $^kD_\theta(P_s)$, or $^kD_\pi(P_s)$, by the other hand. We use the electrostatic field by default and omit it in notations. In the next lines, we show the linear MI models for Drug-Protein Interactions (DPIs).

$$S(DPI_{rs})_{pred} = \sum_{k=0}^{5} a_k \cdot {}^{k}D_t(L_r) + \sum_{k=0}^{5} b_k \cdot {}^{k}D_1(P_s) + c_0 \quad (9)$$

The model deals with the ca1culation of score values (S) to predict the propensity of a set of compounds, to interact ($L_{rs}$ = 1) or not ($L_{rs}$ = 0) with different protein targets. A dummy input variable Affinity Class (AC) codify the affinity; AC = 1 for well known DPIs and AC = 0 otherwise. This variable indicates either high (AC = 1) or low (AC = 0) affinity of the $r^{th}$ drug or protein by the $s^{th}$ target protein. The parameter $S(DPI_{rs})_{pred}$ is the output of the model and a continuous and dimensionless score that give higher values for DPIs and lower values for nDPIs. In the model, $a_k$, $b_y$, $c_k$, and $d_0$ represents the coefficients of the MI function determined using the software STATISTICA 6.0 software package [117]. In all these cases, as well as in all the following models presented here, we can check the Specificity (Sp), Sensitivity (Sn), total Accuracy (Ac), or the Area Under the ROC curve (AUROC) to determine the goodness-of-fit to data in training and external validation series.

### *4.2. MI models of Complex Networks (Nets)*

We can seek a linear function able to discriminate between two classes of pairs of nodes, linked and not linked in a new model network. The data necessary to train the model are obtained from the different systems studied. This data includes two types of pairs of nodes (categorical dependent variable): linked ($L_{ij}$ = 1) and not linked ($L_{ij}$ = 0). The MI function has the following form:

$$S(L_{ij}) = \sum_{k=0}^{5} a_{ik} \cdot {}^{k}C_t(i) + \sum_{k=0}^{5} b_{jk} \cdot {}^{k}C_t(j) + \sum_{k=0}^{5} c_{ijk} \cdot [{}^{k}C_t(i) - {}^{k}C_t(j)] + \sum_{k=0}^{5} d_{ijk} \cdot {}^{k}C_t(i) \cdot {}^{k}C_t(j) + e_0 \quad (10)$$

The continuous dependent variables used are: the node centralities of order k and type t for the two nodes ${}^{k}C_t(i)$, ${}^{k}C_t(j)$ and functions of these node centralities like $[{}^{k}C_t(i) - {}^{k}C_t(j))]$ and ${}^{k}C_t(i) \cdot {}^{k}C_t(j)$. Here we use the symbol ${}^{k}C_t$ instead of $D_t$ (the symbol used in the previous examples). This difference indicates that in the previous examples of MI models we talk in general about descriptors ${}^{k}D_t$ (centralities or not) of a molecular graph. However, in this example we are talking about node centralities ${}^{k}C_t$. Therefore we have Nv = 4·k·t variables that encode information of the pair of nodes ij and its neighbors (placed at a topological distance d = k). The parameters $a_{ik}$, $b_{jk}$, $c_{ijk}$, and $d_{ijk}$ are coefficients for variables and $a_0$ the independent term. $S(L_{ij})$ is the output variable (a real number).

### *4.3. Models with MA Operators*

Let be $S_j$ the output variable of a model used to score the quality of the connectivity pattern $L_{ij}$ between the node $i^{th}$ and all the remnant *(n* -1) nodes in the network. In this sense, $S_j$ is a real valued variable that scores the quality of the connectivity pattern or links (all direct and indirect connections) established between the node $j^{th}$ and the other nodes. The higher is the value of $S_j$ the closer to the correct pattern are the links set for $j^{th}$ in the network as a whole, according to the model. On the other hand, $L_j$ is the input dependent variable. $L_j = 1$ when a node is correctly linked to the rest of the nodes in the network and $L_j = 0$ when a node has a random connectivity model. We can use linear algorithm like Linear Discriminant Analysis (LDA) or a Linear Neural Network (LNN) to fit the coefficients $a_k$, ${}^{g}b_k$, and $c_0$. We can use also a non-linear methods, *e.g.*, Artificial Neural Networks (ANNs) [118]. The linear equation case is:

$$S_j = \sum_{k=0}^{5} a_k \cdot {}^{k}C_t(j) + \sum_{g=0}^{g=N_g} \sum_{k=0}^{5} b_{gk}\left[ {}^{k}C_t(j) - {}^{k}C_t(j)_{g-avg}\right] + c_0$$
$$= \sum_{k=0}^{5} a_k \cdot {}^{k}C_t(j) + \sum_{g=0}^{g=N_g} \sum_{k=0}^{5} b_{gk} \cdot \Delta^{k}C_t(j)_g + c_0 \quad (11)$$

In this equation we can see the coefficients ($a_k$) of the Wiener-Markov centralities used as input $W_k(j)$ and/or the coefficients (${}^{g}b_k$) of different deviation terms constructed with these variables. The deviation terms have the general form $\Delta W_k(j)_g = [W_k(j) - W_k(j)_{g.avg}]$. Where, $W_k(j)_{g.avg}$ is the average value (avg) of $W_k(j)$ for a sub-set or group (g) of nodes of the same graph G ($g \in G$) that obey a given condition. This type of deviation terms resembles the moving average terms used in time series models like in Box- Jenkins' ARIMA models [45]. However, in the present work g may be not only a period or season (laws approved in the same year) but also a biological boundary (metabolic reactions in the same organism) or spatial condition (interactions in the same eco-system); see results section.

## 5. EXAMPLES OF MI MODELS

### *5.1. Markov-Shannon Entropy Models*

Entropy measures are universal parameters useful to codify biologically relevant information in many systems. Kier published probably the first work on the use of Shannon's entropy to ca1culate a structural information parameter (called molecular negentropy) and carry out QSPR studies [119, 120]. Graham *et al.* [121-126] used entropy measures to study the information properties of organic molecules. In any case, Shannon's entropy have been used to describe not only small molecules [120, 127-134] but also protein [135, 136] or DNA sequences [137] as well as protein interaction networks [138]. Mikoláš *et al.* [139] reviewed the use of entropy measures in functional magnetic resonance (fMRl). The software MI ca1culates values of Markov-Shannon entropy for both molecular structures (drugs and target proteins) and nodes centralities in complex networks [84, 92]. Last year [47], we published a paper on the QSPR study of complex molecular systems and social networks using entropy measures and one alignment-free, multi-target, and multi-scale algorithm (see Fig. 1). The procedure is essentially the same than in classic QSPR studies with some variations in each problem. In the following sections, we review some of these MI models for illustrative purposes. The first model was developed to predict the DT-Net of FDA approved drugs. The prediction of DT-Nets is important due to the high cost of the experimental [78, 140, 141]. Here, we have developed a model that takes into account the structure of the drug, the structure of the target, and the information about the drug/target nodes in the studied network (see Fig. 3).

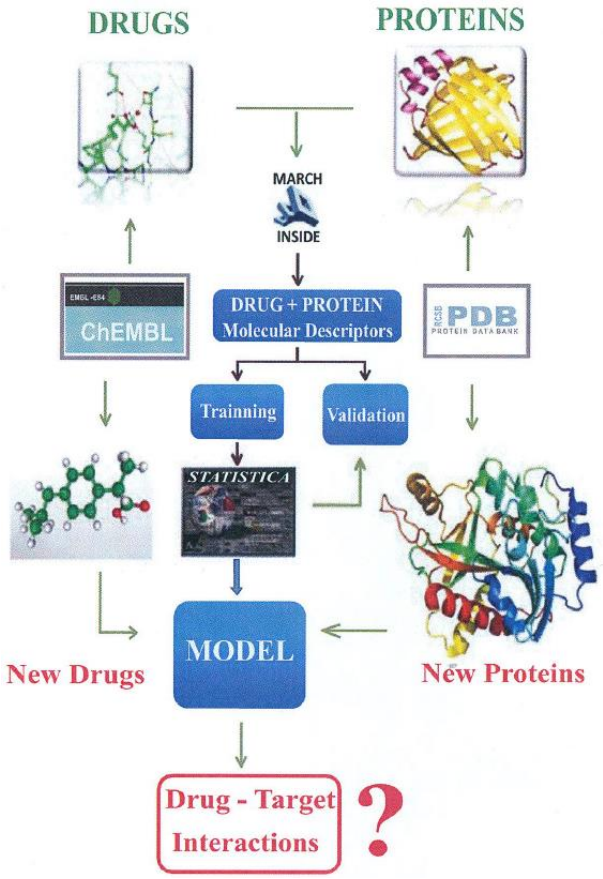



**Fig.(3)**. QSPR analysis of grug-target networks

In this network $L_{rs}$ = 1 if the $r^{th}$ protein ($P_r$) is a target of the $s^{th}$ drug or ligand ($L_s$) in the DrugBank database and $L_{rs}$ = 0 otherwise. The best model found was:

$$S(L_{rs}) = +0.11 \cdot {}^{0}D_{\theta}(L_r) - 0.47 \cdot {}^{4}D_{\theta}({}^{m}P_s)$$
$$-2.19 \cdot {}^{3}C_{\theta}(j)_P - 1.10 \cdot {}^{5}C_{\theta}(j)_L - 1.43 \quad (12)$$
$$n = 2{,}234\ \chi^2 = 2{,}123\ p < 0.001$$

where, ${}^{k}D_{\theta}(L_r)$ and ${}^{k}D_{\theta}({}^{m}P_s)$ are the Markov-Shannon entropy descriptors used to codify the information about the structure of the drug and the protein. Specifically the descriptor of the protein inc1udes amino acids placed only in the middle region (m) of the target proteins (see details about protein descriptors in the previous sections).

In addition, $C_{\theta}(j)_L$ and ${}^{k}C_{\theta}(j)_P$ are centralities of the nodes for the drug/ligand and the target in the DT-Net. This put in evidence the multiscale nature of the model with descriptors for drugs, proteins, and nodes in the DT-Net. The $\chi^2$ = 2,123 statistics corresponds to a p-level < 0.001, which indicates a significant discrimination ratio. The values of Ac, Sn, and Sp were very good for validation and training series, see details in the reference [47].

On the other hand, the study of Metabolic Reaction Networks (MR-Nets) is of great interest in biology because many applications are directly built on the use of cellular metabolism in Biotechnology and Biomedicine [142, 143]. In this sense, computational studies of MR-Nets become very useful [144, 145].

In a recent work [47], we developed a model to predict the correct connectivity patterns in MRNs using as inputs the Markov-Shannon entropy centralities ${}^{k}C_{\theta}(j)$ for nodes in already-known networks. For this analysis, we have used metabolic networks of four model organisms belonging to different domains of the tree of life. These organisms are *Escherichia coli* (EC), *Saccharomyces cerevisiae* (SC), *Caenorhabditis elegans* (CE), and *Oryza saliva* (OS). They cover important branches of the tree of live inc1uding a gram-negative bacterium [146-157], a fungus with industrial importance [158], free-living nematode that has become a popular model for genetic [159-162], and the most widely studied model for cereals [163], respectively. The best MI-Entropy model found was:

$$S(L_{ij}) = 159.16 \cdot {}^{3}C_{\theta}(e_i) - 120.70 \cdot {}^{1}C_{\theta}(p_j) - 95.45 \cdot \left[ {}^{5}C_{\theta}(e_i) - {}^{5}C_{\theta}(p_j) \right] - 0.26 \quad (13)$$
$$n = 74{,}999 \;\; \chi^2 = 26{,}093 \;\; p < 0.001$$

In this equation, $S(L_{ij})$ is a real-valued output variable that scores the propensity of the $i^{th}$ input or educts ($e_i$) (reactant or substrate) to undergo a metabolic transformation into the product ($p_j$), The entropy parameters quantify the information related to middle-long range subsequent metabolic transformations of all the neighbors of the input- output metabolites (k = 5) in the metabolic network. See results in Table 2.

Also, the importance for the human and animal health and therefore for the economy, much attention has been focused on complex network analysis of parasite-host interactions [164]. However, the high experimental difficulty inherent to the *in situ* determination these interactions make the use a computational model a very interesting option. In this work, we used ${}^{k}C_{\theta}(j))$ values to seek a QSPR-like model able to predict HP-Nets. The best model found for the HP- Netwas:

$$S(L_{ij}) = -82.62 \cdot \left[ {}^{5}C_{\theta}(p_i) - {}^{5}C_{\theta}(h_j) \right] - 5.52 \quad (14)$$
$$n = 49{,}218 \;\; \chi^2 = 21{,}728 \;\; p < 0.001$$

In this equation, $S(L_{ij})$ is a real-valued output variable that scores the propensity of the $i^{lh}$ parasite specie ($p_i$) to infect a given host specie ($h_j$). See results in Table 2. Connectivity is also the key to understanding distributed and cooperative brain functions and can be represented by Brain Connectome Networks (BC-Nets) [165].

The eventual impact and success of connectivity databases, however, will require the resolution of several methodological problems that currently limit their use. These problems comprise four main points: (i) objective representation of coordinate- free, parcellation-based data, (ii) assessment of the reliability and precision of individual data, especially in the presence of contradictory reports, (iii) data mining and integration of large sets of partially redundant and contradictory data, and (iv) automatic and reproducible transformation of data between incongruent brain maps [166].

In order to address points (ii) and (iv), we have developed a specific model for the 'collation of connectivity data on the macaque brain' (CoCoMac) database (http://www.cocomac.org). The best model found for this BC-Netwas:

$$S(L_{ij}) = 70.56 \cdot {}^{1}C_{\theta}(a) + 74.51 \cdot {}^{5}C_{\theta}(e) - 1.75$$
$$n = 39{,}536 \quad \chi^2 = 22{,}249 \quad p < 0.001 \tag{15}$$

In this equation, $S(L_{ij})$ is a real-valued output variable that scores the propensity of the $i^{lh}$ cerebral cortex region to undergo co-activation with the $j^{th}$ region in the CoCoMac network. The entropy parameters quantify the information related to the position of the afferent/efferent regions and their direct neighbors (k = 1) in the network. The model showed very good results (see Table 2).

**Table 2**. MI models of complex networks

| Net | Par. | ${}^{k}CBM(j)$ | ${}^{k}C_{\theta}(j)^{k}_{j}$ | ${}^{k}C_{\pi}(j)^{k}_{j}$ | ${}^{k}C_{wc}(j)$ | ${}^{k}C_{\chi}(j)$ |
|---|---|---|---|---|---|---|
| Train | | | | | | |
| MR | Sp | 72.22 | 99.98 | ? | 81.32 | 70.19 |
| | Sn | 71.25 | 87.24 | ? | 73.91 | 70.63 |
| PH | Sp | 87.49 | 95.4 | 87.49 | 95.24 | 90.56 |
| | Sn | 100 | 72.22 | \00 | 73.27 | 92.70 |
| BC | Sp | 84.14 | 92.2 | 98.49 | 88.40 | 75.32 |
| | Sn | 72.70 | 71.2 | 73.30 | 74.64 | 94.69 |
| FE | Sp | 87.14 | 99.2 | 93.21 | 71.49 | 100 |
| | Sn | 72.68 | 70.4 | 12.01 | 71.64 | 89.70 |
| Validation | | | | | | |
| MR | Sp | 72.28 | 99.96 | ? | 81.82 | 71.17 |
| | Sn | 71.24 | 86.91 | ? | 73.81 | 70.89 |
| PH | Sp | 87.67 | 95.5 | 87.67 | 95.43 | 91.00 |
| | Sn | 100 | *12* | 100 | 70.81 | 92.83 |
| BC | Sp | 84.42 | 92.5 | 98.41 | 88.30 | 75.51 |
| | Sn | 71.88 | 70.4 | 71.21 | 73.27 | 94.73 |
| FE | Sp | 87.34 | 99.1 | 93.20 | 71.55 | 100 |
| | Sn | 75.78 | 74.2 | 73.47 | 70.54 | 90.22 |
| Mod. | Ref. | [52] | [47] | [116] | [51] | |

Net - Network. 1- Metabolic Reactions Network (MR-Net), 2 - Parasite-Host Net (PH-Net), 3 - Brain Connectome Net (BC-Net), 4 - Fasciolosis Epidemiology Net (FE-Net). Par. - Parameter: Sp = Specificity and Sn = Sensitivity. Ref. = Reference where the model was published.

Another important problem to be studied with networks is the spreading of diseases. For instance, Fasciolosis is a parasitic infection caused by *Fasciola hepatica* (liver fluke) that has become an important cause of lost productivity in livestock worldwide. It is considered a secondary zoonotic disease until the mid-1990s, human fasciolosis is at present emerging or re-emerging in many countries. In addition, it presents a range of epidemiological characteristics related to a wide diversity of environments [167].

In this sense, the study of geographical spreading of fasciolosis becomes a subject of great interest. In fact, in a recent work we have constructed a Fasciolosis Epidemiology network (FE-Net) to study the landscape spreading of fasciolosis in Galicia (NW Spain) [168]. However, we do not have quantitative criteria on the quality of the network connectivity, and re-sampling of all data to re-evaluate this connectivity in a field study is a hard and expensive task in terms of time and resources.

This situation has prompted us to seek a model in order to assess the quality of the network previously assembled. The best QSPR model found and published in our previous work for the FE-Net was:

$$S(L_{ij}) = -20.23 \cdot {}^{1}C_{\theta}(f_i) + 165.13 \cdot {}^{4}C_{\theta}(f_j) - 0.82$$
$$n = 19{,}671 \quad \chi^2 = 16{,}058 \quad p < 0.001 \tag{16}$$

The entropy used in this equation quantifies information about the connectivity patterns between farms in the network **C**.

As can be seen in the equations described in material s and methods, the connectivity of **C** depends on the spatial coordinates ($x_i$, $y_i$) of the farm ($f_i$), the altitude of the place ($h_i$), and the anti-parasite drug treatment ($Tr_j$) used to prevent Fasciolosis in this farm. Consequently the matrix e quantifies the *a priori* propensity $C_{ij} = 1$ of this disease to spread between farms immediately after treatment depending on geographical conditions.

On the other hand, matrix **L** includes both criteria: (i) the preexistence of a high propensity for disease spreading $C_{ij} = 1$ and (ii) the experimental confirmation $L_{ij} = 1$ of a high Risk Ratio ($RR_{ij}$) of Prevalence After Treatment ($PAT_j$) for this disease in farms. See Fig. (4), published before in one of our papers [52], see al so the section about auto-correlation indices. The QSPR equation developed here was obtained by studying **L** and the model presents good values of Sensitivity (Sn), and Specificity (Sp), see Table 2.

Another MI-Shannon entropy model published in the previous work is useful to study the SL-Net for Spain's law system. The use of network analysis methods in social sciences began in 1930 and today are widely used [169]. However, the application of these methods in legal studies is still at the beginning [170-172]. Network tools may illustrate the interrelation between the different law types and help to understand law consequences in society and its effectiveness or not. We have used the list of the financial laws to construct the network described. The best model found was:

$$S(L_{ij}) = 650.88 \cdot [\,{}^{1}C_{\theta}({}^{c}L_{ti}) - {}^{1}C_{\theta}({}^{c}L_{ti+1})] + 0.12$$
$$n = 33{,}951 \quad \chi^2 = 32{,}942 \quad p < 0.001 \tag{17}$$

where the two parameters in the equation are the entropy parameters that quantify information about the Legal norms (Laws) of type L introduced in the Spanish legal system at time $t_i$ and $_{ti+1}$ with respect to the previous or successive $k^{th}$ norms approved. The model behaves like a time series embedded within a complex network. This is because it predicts the recurrence of the Spanish law system to a financial norm of class c when socio-economical conditions change at time $t_{i+1}$ given that have been used a known class of norm in the past at time $t_i$. The model correctly reconstructed the network of the historic record for the Spanish financial system with high Sp and Sn (Table 2). In Fig. (5), we illustrate the steps used to develop the MI model of this network; which is also a hierarchical time series.

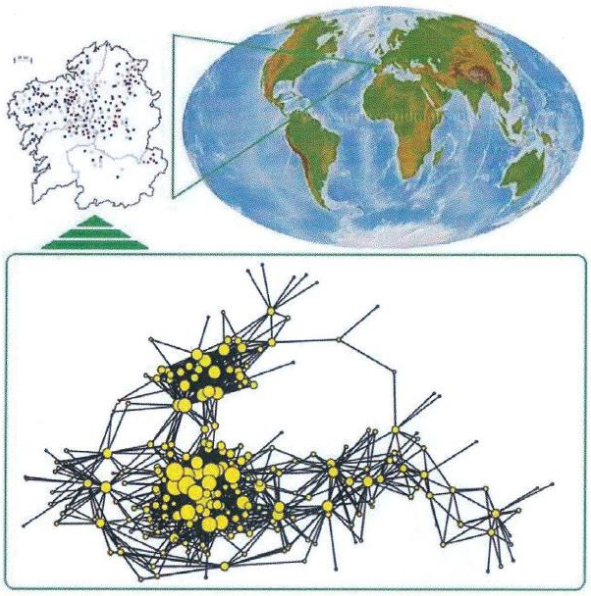


**Fig. (4).** Top left: Geographical map. of Galicia (NW Spain) showing the location of the 275 sampled farms: the status of infection (empty circles: *F. .hepatica* free and filled circles: *F. hepatica* infected) and the treatment administered on each farm are shown (blue: none; red: anthelmintic effective against fluke mature stages and green: a fasciolicide effective against immature and mature stages). Bottom: Fasciolosis landscape-spreading network. The size of each node represents its degree.

The last Ml-Shannon entropy model reported is useful to predict the Network (WT-Net) of Smart Package for World's food industry. Traditionally, the basic functions of packaging have been. classified into 4 categories: protection, communication, convenience, and containment [173]. Smart or Active Packaging is an innovative concept that can be defined as a mode of packaging in which the package, the product, and the environment interact to pro long shelf life or enhance safety or sensory properties, while maintaining the quality of the product [174]. In addition there is a growing concern about foodborne diseases, and many companies are interested in the development of biosensors included in the packages in order to detect the presence of pathogens [173]. In the previous work we studied a large world-trading network (WT-Net) for the current world trade (year 2011) of smart packaging for food industry, interconnecting categories like Country (CU), Company (CO), Product (PR), Food Type (FT), and product use or Packaging Type (PT), see also datasets section. The best model found was:

$$S(L_{ij}) = -2.00 \cdot {}^{1}C_{\theta}(i) - 142.87 \cdot {}^{1}C_{\theta}(j) + 116.65 \cdot {}^{5}C_{\theta}(j) + 0.72 \qquad (18)$$

$$n = 31{,}911 \quad \chi^2 = 19{,}022 \quad p < 0.001$$

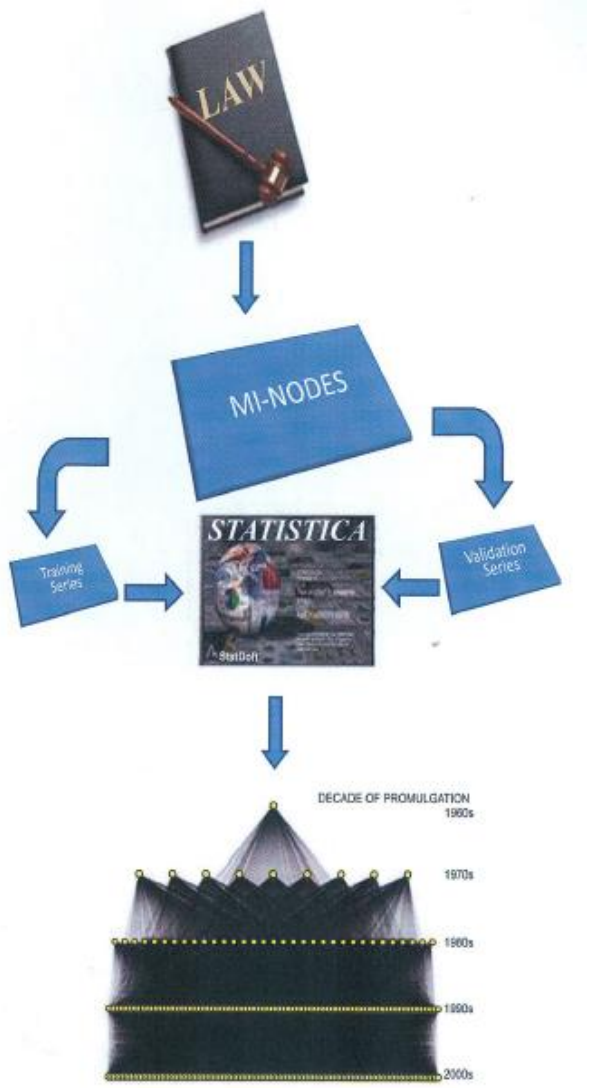



**Fig. (5)**. QSPR analysis of one SL-Net.

The model presents very good values of Sn and Sp (see Table 2). The first parameter quantifies the information referred to the trading relationships of the $i^{lh}$ node with its direct neighbors (k = 1) in the world trade network. The second parameter quantifies the same information for the $j^{th}$ node and its direct neighbors (k = 1). The last parameter quantifies the information referred to middle-long range trading relationships (k = 5) in the trade network between the $j^{th}$ node and its neighbors of any class. In order to use this equation, it is necessary to introduce the values of the centralities for the $i^{lh}$ and $j^{th}$ nodes according to the following hierarchical order in i to j direction: Country (CU) → Company → (CO) → Product (PR) → Packaging Type (PT) → Food Type (FT) if we want to predict the expected success of a given CO to introduce a determined PT in the WT-Net. In Fig. (6), we illustrate the network for a better understanding.

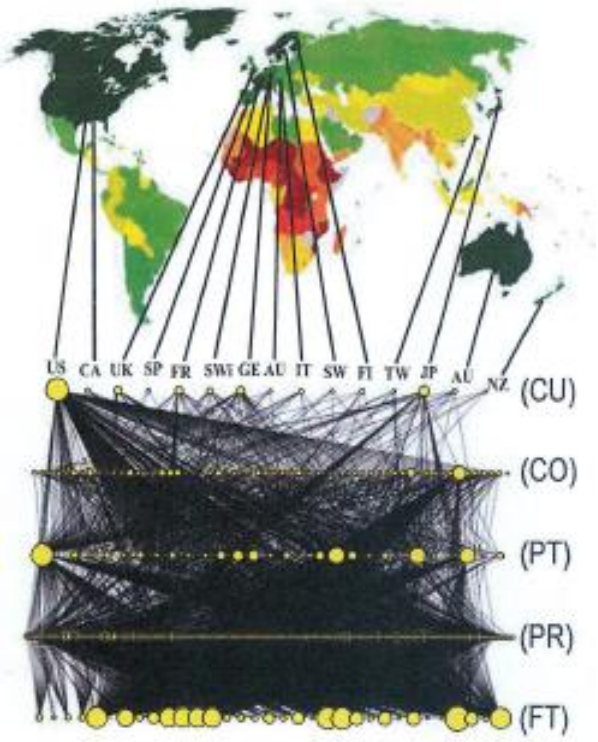



**Fig. (6).** WT -Net of smart packaging for food industry.

### *5.2. Rücker-Markov Centralities Models*

Rücker and Rücker [175, 176] published a series of works about the use of Walk Count (WC) indices, in this sense. In this previous work, it is demonstrated how the complexity of a (molecular) graph can be quantified in terms of the walk counts, extremely easily obtained graph invariants that depend on size, branching, cyclicity, and edge and vertex weights (w). Gutman co-authored another paper with Rücker & Rücker about WCs [177]. They reviewed applications of WCs in theoretical chemistry based on the fact that the (i, j)-entry of the $k^{th}$ power of the adjacency matrix is equal to the number of walks starting at vertex j, ending at vertex j, and having length k. In 2003, the concept was extended by Lukovits and Trinajstié [178] to zero and negative orders. More recently, Bonchev has applied WCs and other TIs to the complexity analysis of yeast proteome network [3]. In a recent work, we introduced the new Rücker-Markov indices ${}^{k}C_{wc}(j)$ [179] and use them to seek QSPR models able to predict of the connectivity of new complex networks. For instance, we used ${}^{k}C_{wc}(j)$ values to seek a QSPR-like model able to predict PH-Nets, the DS-Net of Fasciolosis in Galicia, and the BC-Net reported in CoCoMac experiment. The best models found for each one of these datasets were the following, in this order:

$$\begin{aligned} S(L_{ij}) = & -258.93 \cdot \left[ {}^{1}C_{wc}(p_i) - {}^{1}C_{wc}(h_j) \right] \\ & +283.69 \cdot [ {}^{2}C_{wc}(p_i) - ]^{2}C_{wc}(h_j) \\ & -88.75 \cdot \left[ {}^{4}C_{wc}(p_i) - {}^{4}C_{wc}(h_j) \right] + 0.25 \\ n = 49{,}218 \quad & \chi^2 = 22{,}297 \quad p < 0.001 \end{aligned} \tag{19}$$

$$S(L_{ij}) = 8.34 \cdot \left[ {}^{1}C_{wc}(f_i) - {}^{1}C_{wc}(f_j) \right]$$
$$-2.17 \cdot \left[ {}^{5}C_{wc}(f_i) - {}^{5}C_{wc}(f_j) \right] - 0.56) \quad (20)$$
$$n = 23{,}991 \quad \chi^2 = 1{,}965 \quad p < 0.001$$

$$S(L_{ij}) = 1.92 \cdot {}^{1}C_{wc}(i) + 2.14 \cdot {}^{2}C_{wc}(j) - 1.68$$
$$n = 39{,}070 \quad \chi^2 = 20{,}602 \quad p < 0.001 \quad (21)$$

In these equations, $S(L_{ij})$ is a real-valued output variable that scores the propensities with which the $i^{th}$ parasite specie ($p_i$) infect host specie ($h_j$) , the disease spreads from the $i^{th}$ farm to the $j^{th}$, or the the $i^{th}$ cerebral region to co-activate with the $j^{th}$ region. You can compare the results for those and other models in Table 2.

### *5.3. Broto-Moreau Stochastic Centralities Models*

In the 1980s, Broto & Moreau applied an autocorrelation function to the molecular graph in order to measure the distribution of atomic properties on the molecular topology. This measure was called Autocorrelation of Topological Structure (ATS) or Broto-Moreau autocorrelation indices (BMis) [180-182]. The idea of ATS has been re-formulated in different ways in order to incorporate more information. Moro studied electrostatic potential surface properties [183], Caballero and Fernández [184-187] carry out QSPR in proteins. Some ATS models have been implemented in web servers such as IUPforest-L [188] and PROFEAT [189]. We implemented them in the software S2SNet (Sequence to Star Networks) [190], to calculate ATS indices for mass spectra signals of proteins, 1D NMR signals, IR spectra, time series data, texts and any other type of string data. In a recent work we studied similar datasets than in the two previous examples but using the MI autocorrelation centrality values ${}^{k}C_{BM}(j)$ [191]. The best model for the MR-Nets of the organisms EC, SC, CE, and OS, PH-Nets, BC-Net of macaque visual cortex, and DS-Net for Fasciolosis in Galicia are the following, see also Table 2.

$$S(L_{ij}) = -0.73 + 23.44 \cdot {}^{5}C_{BM}(e_i)$$
$$-5.59 \cdot \left[ {}^{3}C_{BM}(e_i) - {}^{3}C_{BM}(p_j) \right] \quad (22)$$
$$n = 74{,}999 \quad \chi^2 = 20{,}143 \quad p < 0.001$$

$$S(L_{ij}) = 4.59 \cdot \left[ {}^{2}C_{BM}(p_i) - {}^{2}C_{BM}(h_j) \right] + 0.21$$
$$n = 49{,}218 \quad \chi^2 = 15{,}801 \quad p < 0.001 \quad (23)$$

$$S(L_{ij}) = 12.74 \cdot [ {}^{1}C_{BM}(i) - {}^{1}C_{BM}(j)] - 0.80$$
$$n = 24{,}956 \quad \chi^2 = 9{,}422 \quad p < 0.001 \quad (24)$$

$$S(L_{ij}) = -0.07 - 11.50 \cdot {}^{3}C_{BM}(f_i)$$
$$-18.26 \cdot \left[ {}^{1}C_{BM}(f_i) - {}^{1}C_{BM}(f_j) \right] \quad (25)$$
$$n = 23{,}377 \quad \chi^2 = 3{,}897 \quad p < 0.001$$

### *5.4. Wiener-Markov Centralities Models*

In 1947, Wiener published an article entitled *Structural determination of paraffin boiling points* [192]. In this work, it is proposed that organic compounds, as well as all their physical properties, depend functionally upon the number, kind, and structural arrangement of the atoms in the molecule [193-195]. Hosoya coined one term of Wiener's equation in 1971 as the Z index [196-198].

The Wiener index (W) index was independently proposed in 1959 by Harary in the context of sociometry, with the name *total status of a graph* [199] as well as in 1975 by Rouvray and Crafford [200]. In any case, W index or path number is calculated as the half sum of all the elements $d_{ij}$ of the distance matrix (D). More distant atom pairs make larger contribution to W than adjacent atom pairs:

$$W = \frac{1}{2} \cdot \sum_{i=1}^{D} \sum_{j=1}^{D} d_{ij} \tag{26}$$

In a very recent work [53], we used Markov-Wiener centralities ${}^{k}C_{w}(j)$ to predict correct connectivity patterns of nodes in MR-Nets of 43 organisms using MIANN models (acronym formed by MI and ANN)[9]. In Table 3, we depict the classic parameters and the average values of ${}^{k}C_{w}(j)$ for the full MR-Nets of many organisms. These average values are the inputs used to characterize the organisms with the MI method in the predictive MIANN models. After that, we tested different MIANN models using as inputs the values of ${}^{k}C_{w}(j)$ and with linear (LNN) and non-linear (ANN) topologies in of the ANN.

In Table 4, we can see that the best MIANN model found presents very good values of Accuracy, Sensitivity, and Specificity for the recognition of links both in training and external validation series. The models were obtained using as input 15 descriptors: 5 Markov- Wiener centralities ${}^{k}C_{w}(j)$, 5 MA values denoted as ${}^{k}C_{w}(j)_{g.avg}$ and 5 deviation terms $\Delta\, {}^{k}C_{w}(j)_{g}$. Multilayer Perceptrons (MLP) [201] method fails to generate good prediction models, since it presents values of Specificity and Sensitivity close to 50%. On the other hand, the LNN based on 15 descriptors (LNN 15:15-1:1) is able to classify correctly a 78.1 % of the cases, with a sensitivity of 77.9% and a specificity of 77.6%. The LNN is equivalent to a LDA equation, the simplest type of classification model.

We also developed a MIANN-Wiener models of BI-Nets published in IWDB. The results are presented in Table 4. We obtained the best classification model for IWDB with the MLP classifier based on 13 input descriptors and 13 neurons in the hidden layer (MLP 13:13-13-1:1). This model can classify 91.1 % of the nodes with a sensitivity of 90.5% and specificity of 88.8%. Unlike the case of the MRNs, the LNN is not able to classify the nodes in the BI-Net with accuracy (<67%). Thus, the BI-Nets contain complex information for the classification of the connectivity between nodes. The IWDBNs need complex classifiers such as MLPs in comparison with the MR-Nets that can be processed by using the simpler LNNs.

The Fig. (7) depicts one illustration of the IWDB. Last, we reported a MIANN-Wiener model for SL-Net of Spain's financial law system. These MIANN models behave like time series embedded within a complex network. The model predicts the recurrence of the Spanish law system to a financial norm of class e when the socio-economical conditions change at time $t_{i+1}$ given that have been used a known class of norm in the past at time $t_i$. The best model correctly re-constructed the network of the historic record for the Spanish financial system with high Sp and Sn (see Table 4).

In this case, there is not a clear difference between the two models studied (LNN and MLP). In this situation, we can apply the Occam's razor and choose the LNN model, which is the simplest.

**Table 3**. Classic parameters *vs* Average values ${}^{k}C_{w}(j)_{org.avg}$ of metabolic networks of different organisms.

| Organism | Classic Parameters ofMRNs | | | | | | | | Markov- Wiener Centralities | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SYMBOL | N | $L_{in}$ | $L_{out}$ | R | E | $g_{in}$ | g.,,, | D | k=1 | k=2 | k=3 | k=4 | k=5 |
| | | | | | | | | | | | | | |
| AA | 419 | 1278 | 1249 | 401 | 285 | 2.1 | 2.2 | 3.3 | 0.87 | 1.08 | 1.26 | 1.44 | 1.57 |
| AB | 395 | 1202 | 1166 | 380 | 271 | 2.1 | 2.2 | 3.2 | 0.88 | 1.07 | 1.24 | 1.43 | 1.56 |
| AG | 496 | 1527 | 1484 | 486 | 299 | 2.2 | 2.2 | 3.5 | 0.85 | 1.09 | 1.29 | 1.48 | 1.61 |
| AP | 204 | 588 | 575 | 178 | 135 | 2.2 | 2.2 | 3.2 | 0.95 | 1.11 | 1.25 | 1.46 | 1.6 |
| AT | 302 | 804 | 789 | 250 | 185 | 2.1 | 2.3 | 3.5 | 0.89 | 1.12 | 1.3 | 1.48 | 1.62 |
| SS | 187 | 442 | 438 | 140 | 106 | 2.3 | 2.4 | 3 | 0.8 | 0.99 | 1.18 | 1.37 | 1.49 |
| SS | 785 | 2794 | 2741 | 916 | 516 | 2.2 | 2.1 | 3.3 | 0.8 | 1.09 | 1.3 | 1.52 | 1.65 |
| CA | 494 | 1624 | 1578 | 511 | 344 | 2.1 | 2.2 | 3.3 | 0.83 | 1.08 | 1.28 | 1.46 | 1.59 |
| CE | 462 | 1446 | 1418 | 450 | 295 | 2.1 | 2.2 | 3.3 | 0.9 | 1.12 | 1.32 | 1.51 | 1.65 |
| *Cl* | 380 | 1142 | 1115 | 359 | 254 | 2.1 | 2.3 | 3.2 | 0.88 | 1.09 | 1.27 | 1.45 | 1.58 |
| CL | 389 | 1097 | 1062 | 333 | 231 | 2.1 | 2.2 | 3.3 | 0.88 | 1.1 | 1.3 | 1.51 | 1.63 |
| CQ | 194 | 401 | 391 | 134 | 84 | 2.2 | 2.3 | 3.4 | 0.99 | 1.14 | 1.27 | 1.47 | 1.62 |
| CT | 215 | 479 | 462 | 158 | 94 | 2.2 | 2.4 | 3.5 | 0.9 | 1.06 | 1.22 | 1.38 | 1.5 |
| CY | 546 | 1782 | 1746 | 570 | 370 | 2 | 2.2 | 3.3 | 0.88 | 1.13 | 1.33 | 1.56 | 1.68 |
| DR | 815 | 2870 | 2811 | 965 | 557 | 2.2 | 2.1 | 3.3 | 0.89 | 1.12 | 1.31 | 1.52 | 1.65 |
| EC | 778 | 2904 | 2859 | 968 | 570 | 2.2 | 2.1 | 3.2 | 0.79 | 1.03 | 1.24 | 1.44 | 1.57 |
| EF | 386 | 1244 | 1218 | 382 | 281 | 2.1 | 2.2 | 3.1 | 0.81 | 1.04 | 1.24 | 1.42 | 1.55 |
| EN | 383 | 1095 | 1081 | 339 | 254 | 2.1 | 2.2 | 3.3 | 0.89 | 1.11 | 1.31 | 1.5 | 1.65 |
| HI | 526 | 1773 | 1746 | 597 | 361 | 2.1 | 2.3 | 3.2 | 0.77 | 1.05 | 1.26 | 1.48 | 1.59 |
| HP | 375 | 1181 | 1144 | 375 | 246 | 2 | 2.3 | 3.3 | 0.89 | 1.11 | 1.3 | 1.5 | 1.62 |
| MS | 429 | 1247 | 1221 | 391 | 282 | 2.2 | 2.2 | 3.2 | 0.87 | 1.09 | 1.27 | 1.46 | 1.6 |
| MG | 209 | 535 | 525 | 196 | 85 | 2.4 | 2.2 | 3.5 | 0.96 | 1.14 | 1.26 | 1.38 | 1.48 |
| *Ml* | 424 | 1317 | 1272 | 415 | 264 | 2.2 | 2.3 | 3.5 | 0.88 | 1.11 | 1.29 | 1.47 | 1.6 |
| ML | 422 | 1271 | 1244 | 402 | 282 | 2.2 | 2.2 | 3.2 | 0.83 | 1.06 | 1.25 | 1.44 | 1.58 |
| MP | 178 | 470 | 466 | 154 | 88 | 2.3 | 2.2 | 3.2 | 0.91 | 1.11 | 1.29 | 1.46 | 1.59 |
| MT | 587 | 1862 | 1823 | 589 | 358 | 2 | 2.2 | 3.3 | 0.88 | 1.12 | 1.32 | 1.55 | 1.67 |
| NG | 406 | 1298 | 1270 | 413 | 285 | 2.1 | 2.2 | 3.2 | 0.85 | 1.06 | 1.24 | 1.42 | 1.56 |
| NM | 381 | 1212 | 1181 | 380 | 271 | 2.2 | 2.2 | 3.2 | 0.86 | 1.08 | 1.27 | 1.45 | 1.59 |
| OS | 292 | 763 | 751 | 238 | 178 | 2.1 | 2.3 | 3.5 | 0.93 | 1.19 | 1.39 | 1.57 | 1.71 |
| PA | 734 | 2453 | 2398 | 799 | 490 | 2.1 | 2.2 | 3.3 | 0.87 | 1.1 | 1.29 | 1.52 | 1.65 |
| PF | 316 | 901 | 867 | 283 | 191 | 2 | 2.3 | 3.4 | 0.93 | 1.14 | 1.33 | 1.5 | 1.65 |
| PG | 424 | 1192 | 1156 | 374 | 254 | 2.2 | 2.2 | 3.3 | 0.85 | 1.06 | 1.24 | 1.41 | 1.54 |
| PH | 323 | 914 | 882 | 288 | 196 | 2 | 2.2 | 3.4 | 0.92 | 1.12 | 1.31 | 1.49 | 1.63 |
| SC | 561 | 1934 | 1889 | 596 | 402 | 2 | 2.2 | 3.3 | 0.88 | 1.11 | 1.31 | 1.54 | 1.68 |
| ST | 403 | 1300 | 1277 | 404 | 280 | 2.1 | 2.2 | 3.1 | 0.89 | 1.08 | 1.24 | 1.44 | 1.57 |
| TH | 430 | 1374 | 1331 | 428 | 280 | 2.2 | 2.2 | 3.4 | 0.89 | 1.13 | 1.33 | 1.52 | 1.65 |
| TM | 338 | 1004 | 976 | 302 | 223 | 2.1 | 2.2 | 3.2 | 0.88 | 1.09 | 1.28 | 1.47 | 1.6 |
| | | | | | | | | | | | | | |

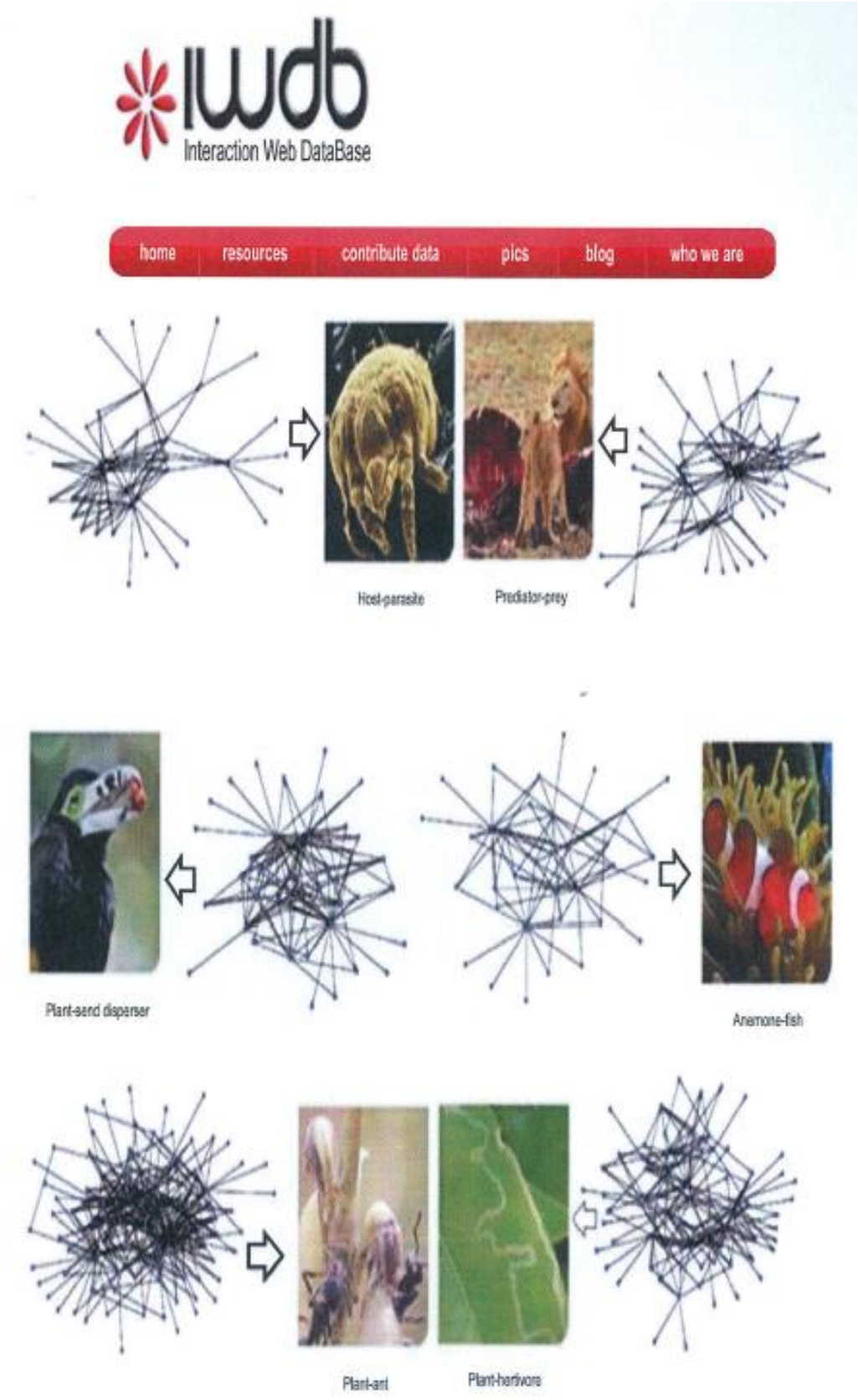
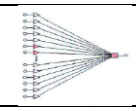
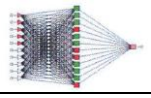
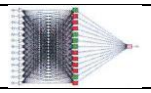
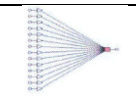
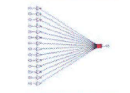
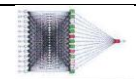
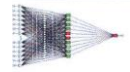
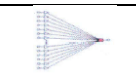



**Fig.(7).** IWDB *vs* BI-Nets

**Table 4**. Some QSPR models of MR-Nets, BI-Nets, and SI-Nets

| Dataset and Model Used | ANN | Li | Train | | | Pr. | Validation | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Li=1 | Li=0 | % | | % | Li=1 | Li=0 |
| | | | | | | | | | |
| Markov-Wiener models of MR-Nets of >40 organisms | LNN 15:15-1:1 | Li=1 | 7276 | 1985 | 78.1 | Sn | 77.9 | 21917 | 6156 |
| |  | Li=0 | 2044 | 7066 | 78.1 | Sp | 77.6 | 6227 | 21329 |
| | MLP 2:2-11-1:1 | Li=1 | 4669 | 4559 | 50.1 | Sn | 49.7 | 13990 | 13856 |
| |  | Li=0 | 4651 | 4492 | 49.6 | Sp | 49.6 | 14154 | 13629 |
| Markov-Wiener models of BI-Nets of >70 ecosystems | MLP 13:13-13-1:1 | Li=1 | 4570 | 547 | 91.1 | Sn | 90.5 | 1363 | |
| |  | Li=0 | 449 | 4346 | 88.8 | Sp | 88.1 | 143 | |
| | LNN 14:14-1:1 | Li=1 | 3326 | 1710 | 66.3 | Sn | 66.1 | 995 | 603 |
| |  | Li=0 | 1693 | 3183 | 65.1 | Sp | 63.0 | 511 | 1028 |
| Markov-Wiener models of SL-Net for Spain's Financial Law system | LNN 14:14-1:1 | Li=1 | 125 | 41 | 86.2 | Sn | 87.4 | 370 | 156 |
| |  | Li=0 | 18 | 298 | 85.4 | Sp | 87.9 | 59 | 914 |
| | MLP 14:14-14:1 | Li=1 | 119 | 54 | 85.3 | Sn | 83.2 | 366 | 129 |
| |  | Li=0 | 24 | 285 | 87.9 | Sp | 84.1 | 63 | 941 |
| Markov-Balaban model for SL-Nets dataset of 5KCNs of USSC | MLP 18:18-10-1:1 | Li=1 | 81225 | 51008 | 82.49 | Sn | 82.76 | 26985 | 17014 |
| |  | Li=0 | 16917 | 243415 | 82.66 | Sp | 82.7 | 5728 | 81128 |
| | LNN 18:18-1:1 | Li=1 | 77871 | 60826 | 79.33 | Sn | 79.35 | 25950 | 20284 |
| |  | Li=0 | 20271 | 233597 | 79.33 | Sp | 79.3 | 6763 | 77858 |

### *5.5. Markov-Balaban Index Models*

Prof. Alexandru T Balaban introduced one of the more famous TIs that have been widely-known as the Balaban's J index [202]. Balabau's J index have been used in many chemo-informatics to quantify structural information and include parameters like q = number of edges in the molecular graph, μ = (q - n + 1) = the cyclomatic number of the molecular graph, n = number of atoms in the molecular graph, and $S_i$ = distance sums calculated as the sums over the rows or columns of the topological distance matrix **D** of the graph G. The formula ofthis classic TI is:

$$J(G) = \frac{q}{\mu + 1} \cdot \sum_{edges}^{q} \left(S_i \cdot S_j\right)^{-1/2} \tag{27}$$

Many applications of Balaban's J index deal with drug discovery; in particular the prediction of drugs with higher biological activity and/or low toxicity [203-210]. J index is useful as input for both linear and non-linear models like ANNs [211, 212]. J index have been used also to compare graphs or analyze combinatorial libraries and some authors have reported new generalizations of this index to create other TIs (called Balaban type parameters) [213-215].

For instance, Randić and Pompe [216]; reported the variable Balaban J index and the "reversed" Balaban index 1/J as well as a novel index 1/JJ derived from J and 1/J. In another very recent work [217], we introduced new Balaban type indices called the Markov-Balaban $^{k}C_{J}(j)$ centralities of order $k^{th}$ for the $j^{th}$ node in a complex network (see Table 1). In this previous work, we also used multiscale MA operators to calculate deviation terms with the general form $\Delta\ ^{k}C_{J}(j)_{g} = [^{k}C_{J}(j) - {}_{k}C_{J}(j)_{g.avg}]$. Where, $TI_{k}(j)_{g.avg}$ is the average value (avg) of $TI_{k}(j)$ for a sub-set or group (g) of nodes of the same graph G ($g \in G$) that obey a given condition. We studied some collections of complex systems like MR-Nets of >40 organisms, BI-Nets of >70 ecological systems, and the SL- Net for all citations to cases of the US Supreme Court (USSC). In this case, g is not only a period (laws approved in the same year), a biological boundary (metabolic reactions in the same organism), or spatial condition (interactions in the same eco-system), but also cases citing the same USSC case. In the last problem we used a SL-Net constructed by Fowler *et al.* [218] with all cases that cite decisions of this court from 1791 to 2005. In the SL-Net of the USSC node represented a legal cases interconnected by arcs to express that the case $j^{th}$ cites the $i^{th}$ case before it (precedent). We constructed in total 43 sub-networks and calculated their $_{k}C_{J}(j)$ values and developed LNN and ANN models to predict them obtaining good results (see Table 4).

## CONCLUSION

In this work, we reviewed the recent results published about the development of MI models. We noted an evolution of MI from a simple one-target chemo-informatics algorithm for series of analogues compounds to models that are more powerful. In this sense, we illustrated the uses of the MI algorithm to solve QSPR problems in Drug- Target, Parasite- Host, Disease Spreading, Brain connectome, and Social- Legal networks. We also showed the different parameters implemented in the MI algorithm to characterize complex networks combining both classic TIs and Markov chains theory. We hope that this review may serve as inspiration to those interested on flexible, fast, and theoretically simple models for the prediction of structure-property relationships in complex systems.

## REFERENCES


Strogatz SH. Exploring complex networks. Nature 2001; 410: 268-76.

Estrada E. Universality in protein residue networks. Biophys J 2010; 98: 890-900.

Bonchev D. Complexity analysis ofyeast proteome network. Chem Biodivers 2004; 1: 312-26.

Gonzalez-Diaz H. Network topological in dices, drug metabolism, and distribution. Curr Drug Metab 2010; 11: 283-4.

Jeong H, Tombor B, Albert R, *et al.* The large-scale organization of metabolic networks. Nature 2000; 407: 651-4.

Estrada E. Returnability as a criterion of disequilibrium in atmospheric reactions network. J Math Chem 2012; 50:1363-72.

Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998; 393: 440-2.

Fornito A, Zalesky A, Pantelis C, *et al.* Schizophrenia, neuroimaging and connectomics. Neuroimage 2012; 62: 2296-314.

González-Díaz H, Arrasate S, Sotomayor N, *et al.* MIANN Models in Medicinal, Physical and Organic Chemistry. Curr Top Med Chem 2013; 13: 619-41.

Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. PLoS One 2010; 5: e12948.

Hu H, Myers S, Colizza V, *et al.* WiFi networks and malware epidemiology. Proc Natl Acad Sci USA 2009; 106: 1318-1323.

Chen Z, Ji C. Spatial-temporal modeling of malware propagation in networks. IEEE Trans Neural Netw 2005; 16: 1291-303.

Krause J, Ruxton GD, Krause S. Swarm intelligence in animals and humans, Trends Ecol Evo12010; 25: 28-34.
Apicella CL, Marlowe FW, Fowler JH, *et al.* Social networks and cooperation in hunter-gatherers. Nature 2012; 481: 497-501.
Fowler JH, Johnson TR, II JFS, *et al.* Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court 2007.
Bornholdt S, Schuster HG. Handbook of Graphs and Complex Networks: From the Genome to the Internet. Wheinheirn: WILEY- VCH GmbH & CO. KGa., 2003.
Boccaletti S, Latora V, Moreno Y, *et al.* Complex networks: Structure and dynamics, Physics Reports 2006; 424: 175-308.
Dehmer M, Emmert-Streib F. Analysis of complex networks: from biology to linguistics. Wheinheirn: Wiley-Blackwell, 2009, 462.
Strogatz SH. Complex systems: Romanesque networks. Nature 2005; 433(7024): 365-6.
Ratti C, Sobolevsky S, Calabrese F, *et al.* Redrawing the map of Great Britain from a network of human interactions. PLoS One 2010; 5: e14248.
Newman ME, Watts DJ, Strogatz SH. Random graph models of social networks. Proc Natl Acad Sci U S A. United States, 2002, 2566-2572.
Newman M. The Structure and Function of Complex Networks. SIAM Review 2003:167-256.
Thomas S, Bonchev D. A survey of current software for network analysis in molecular biology. Hum Genomics 2010; 4: 353-60.
Bonchev D, Buck GA. From molecular to biological structure and back. J Chem InfModel2007; 47: 909-17.
Bonchev D, Rouvray DH. Complexity in Chemistry, Biology, and Ecology. New York: Springer Science+Business Media, Inc, 2005.
Bonchev D. Complexity analysis of yeast proteome network. Chem Biodivers 2004; 1: 312-26.
Bonchev D. On the complexity of directed biological networks. SAR QSAR Environ Res 2003; 14: 199-214.
González-Díaz H. QSAR and complex networks in pharmaceutical design, microbiology, parasitology, toxicology, cancer, and neurosciences. Curr Pharm Des 2010; 16: 2598-600.
Vina D, Uriarte E, Orallo F, *et al.* Alignment-Free Prediction of a Drug-Target Complex Network Based on Parameters of Drug Connectivity and Protein Sequence of Receptors. Mol Pharm 2009; 6: 825-35.
Duardo-Sanchez A, Patlewicz G, González-Díaz H. A Review of Network Topological Indices from Chem-Bioinformatics to Legal Sciences and back. Curr Bioinfo 2011; 6: 53-70.
Tenazinha N, Vinga S. A survey on methods for modeling and analyzing integrated biological networks. IEEE/ACM Trans Comput Biol Bioinforrn 2011; 8: 943-58.
Puzyn T, Leszczynski J, Cronin MTD. Recent Advances in QSAR Studies: Methods and applications. In: Leszczynski J. (ed) Challenges and advances in computational chemistry and physics. Springer 2010, 423.
Estrada E, Rodríguez-Velázquez JA. Subgraph centrality in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 2005; 71: 056103.
Junker BH, Koschutzki D, Schreiber F. Exploration of biological network centralities with CentiBiN. BMC Bioinformatics 2006; 7: 219.
González-Diaz H, González-Díaz Y, Santana L, *et al.* Proteomics, networks and connectivity indices. Proteomics 2008; 8: 750-78.
Kuhn M, Szklarczyk D, Franceschini A, *et al.* STITCH 3: zooming in on protein-chemical interactions. Nucleic Acids Res 2012; 40: D876-80.
Kuhn M, Szklarczyk D, Franceschini A, *et al.* STITCH 2: an interaction network database for small molecules and proteins. Nucleic Acids Res 2010; 38: D552-6.
Kuhn M, von Mering C, Campillos M, *et al.* STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 2008; 36: D684-8.
Zhu F, Shi Z, Qin C, *et al.* Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. Nucleic Acids Res 2012; 40: DI128-36.
Hecker N, Ahmed J, von Eichborn J, *et al.* Super Target goes quantitative: update on drug-target interactions. Nucleic Acids Res 2012; 40: D1113-7.
Gunther S, Kuhn M, Dunkel M, *et al.* Super Target and Matador: resources for exploring drug-target relationships. Nucleic Acids Res 2008; 36: D919-22.
Gaulton A, Bellis LJ, Bento AP, *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Research 2012; 40: DI 100-07.
Mok NY, Brenk R. Mining the ChEMBL database: an efficient chemo informatics workflow for assembling an ion channel-focused screening library. J Chem Inf Model 20 11; 51: 2449-54.
González-Díaz H, Munteanu CR. Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks. Kerala, India: Transworld Research Network 2010, 001-212.
Box GEP, Jenkins GM. Time series analysis. Holden-Day, 1970.
Riera-Fernández P, Munteanu CR, Martín-Romalde R, *et al.* Markov-Randić Indices for QSPR Re-Evaluation of Metabolic, Parasite-Host, Fasciolosis Spreading, Brain Cortex and Legal- Social Complex Networks. Curr Bioinfo 2013; 8: 401-15.

Riera-Fernández P, Munteanu CR, Escobar M, *et al.* New Markov- Shannon Entropy models to assess connectivity quality in complex networks: from molecular to cellular pathway, Parasite-Host, Neural, Industry, and Legal-Social networks. J Theor Biol 2012; 293: 174-88.
Riera-Fernández P, Martin-Romalde R, Prado-Prado FJ, *et al.* From QSAR models of drugs to complex networks: state-of-art review and introduction of new Markov-spectral moments indices. Curr Top Med Chem 2012; 12: 927-60.
Riera-Fernández P, Munteanu CR, Pedreira-Souto N, *et al.* Definition of Markov-Harary Invariants and Review of Classic Topological Indices and Databases in Biology, Parasitology, Technology, and Social-Legal Networks. Curr Bioinfo 20 11; 6: 94- 121.
Riera-Fernández P, Munteanu CR, Dorado J, *et al.* From Chemical Graphs in Computer-Aided Drug Design to General Markov- Galvez Indices of Drug-Target, Proteome, Drug-Parasitic Disease, Technological, and Social-Legal Networks. Curr Comput Aided Drug Des 2011; 7: 315-37.
González-Díaz H, Riera-Fernández P, Pazos A, *et al.* The Rucker- Markov invariants of complex Bio-Systems: Applications in Parasitology and Neuroinformatics. Biosystems 2013; 111: 199- 207.
González-Díaz H, Riera-Fernandez P. New Markov- Autocorrelation Indices for Re-evaluation of Links in Chemical and Biological Complex Networks used in Metabolomics, Parasitology, Neurosciences, and Epidemiology. J Chem lnf Model 2012; 52: 3331-40.
Duardo-Sánchez A, Munteanu CR, Riera-Fernández P, *et al.* Modelling complex metabolic reactions, ecological systems, and financial-legal networks with MlANN models based on Markov- Wiener centralities. Journal of Chemical Information and Modelling 2013 :submitted.
Patil KR, McHardy AC. Alignment-free genome tree inference by learning group-specific distance metrics. Genome Biol Evol 2013.
Cozzetto D, Tramontano A. Relationship between multiple sequence alignments and quality of protein comparative models. Proteins 2005; 58: 151-7.
Iliopoulos I, Tsoka S, Andrade MA, *et al.* Evaluation of annotation strategies using an entire genome sequence. Bioinformatics 2003; 19: 717-26.
Minneci F, Piovesan D, Cozzetto D, *et al.* FFPred 2.0: improved homology-independent prediction of gene ontology terms for eukaryotic protein sequences. PLoS One 2013; 8: e63754.
Wood DE, Lin H, Levy-Moonshine A, *et al.* Thousands of missed genes found in bacterial genomes and their analysis with COMBREX. Biol Direct 2012; 7: 37.
Lynch M. Intron evolution as a population-genetic process. Proc Natl Acad Sci U S A 2002; 99: 6118-23.
Vinga S, Almeida J. Alignment-free sequence comparison-a review. Bioinformatics 2003; 19: 513-23.
Chou KC. Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology. Curr Proteomics 2009; 6: 262-74.
Shen HB, Yang J, Chou KC. Methodology development for predicting subcellular localization and other attributes of proteins. Expert Rev Proteomics 2007; 4: 453-63.
Randic M, Zupan J, Balaban A T, *et al.* Graphical representation of proteins. Chem Rev 2011; 111: 790-862.
González-Díaz H, Dea-Ayuela MA, Pérez-Montoto LG, *et al.* QSAR for RNases and theoretic-experimental study of molecular diversity on peptide mass fingerprints of a new Leishrnania infantum protein. Molecular Diversity 2010; 14: 349-69.
Aguero-Chapin G, Varona-Santos J, de la Riva GA, *et al.* Alignment-Free Prediction of Polygalacturonases with Pseudofolding Topological Indices: Experimental Isolation from Coffea arabica and Prediction of a New Sequence. J Proteome Res 2009; 8: 2122-8.
González-Díaz H, Pérez-Montoto LG, Duardo-Sánchez A, *et al.* Generalized lattice graphs for 2D-visualization of biological information. J Theor Biol 2009; 261: 136-47.
Pérez-Belfo A, Munteanu CR, Ubeira FM, *et al.* Alignment-free prediction of mycobacterial DNA promoters based on pseudo-folding lattice network or star-graph topological indices. J Theor BioI2009;256:458-66.
Agueero-Chapin G, González-Díaz H, de la Riva G, *et al.* MMM- QSAR recognition of ribonuc1eases without alignment: Comparison with an HMM model and isolation from Schizosaccharomyces pombe, prediction, and experimental assay of a new sequence. J Chem lnfModel2008; 48: 434-48.
Dea-Ayuela MA, Pérez-Castillo Y, Meneses-Marcel A, *et al.* HP- Lattice QSAR for dynein proteins: Experimental proteomics (2D- electrophoresis, mass spectrometry) and theoretic study of a Leisbmania infantum sequence. Bioorg Med Chem 2008; 16: 7770- 6.
González-Díaz H, Agueero-Chapin G, Varona J, *et al.* 2D-RNA- coupling numbers: A new computational chemistry approach to link secondary structure topology with biological function. J Comput Chem 2007; 28: 1049-56.
González-Díaz H. Network Topological Indices, Drug Metabolism, and Distribution. Curr Drug Metab 2010; 11: 283-4.
González-Díaz H, Romaris F, Duardo-Sánchez A, *et al.* Predicting Drugs and Proteins in Parasite Infections with Topological Indices of Complex Networks: Theoretical Backgrounds, Applications and Legal Issues. Curr Pharm Des 2010; 16: 2737-64.
González-Díaz H. QSAR and Complex Networks In Pharrnaceutical Design, Microbiology, Parasitology, Toxicology, Cancer and Neurosciences. Curr Pharm Des 2010; 16: U2598- U2524.

González-Díaz H, Munteanu CR. Topological indices for medicinal chemistry, biology, parasitology, neurological and social networks, Kerala: Transworld Research Network 2010.
Hu Y, Bajorath J. Molecular scaffolds with high propensity to form multi-target activity cliffs. J Chem InfModel 2010; 50: 500-10.
Erhan D, L'Heureux P J, Yue SY, *et al.* Collaborative filtering on a family of biological targets, J Chem InfModel 2006; 46: 626-35.
Namasivayam V, Hu Y, Balfer J, *et al.* Classification of compounds with distinct or overlapping multi-target activities and diverse molecular mechanisms using emerging chemical patterns. J Chem InfModel 2013; 53: 1272-81.
Yildirim MA, Goh KI, Cusick ME, *et al.* Drug-target network. Nat Biotech 2007; 25: 1119-26.
Yamanishi Y, Araki M, Gutteridge A, *et al.* Prediction of drug- target interaction networks from the integration of chemical and genomic spaces. Bioinforrnatics 2008; 24: i232-40.
Csermely P, Korcsmaros T, Kiss HJ, *et al.* Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. Pharmacol Ther 2013; 138: 333-408.
Prado-Prado F, García-Mera X, Escobar M, *et al.* 3D MI- DRAGON: New Model for the Reconstruction of US FDA Drug- Target Network and Theoretical-Experimental Studies of Inhibitors of Rasagiline Derivatives for AChE. Curr Top Med Chem 2012; 12: 1843-65.
Prado-Prado F, García-Mera X, Escobar M, *et al.* 2D MI- DRAGON: a new predictor for protein-ligands interactions and theoretic-experimental studies of US FDA drug-target network, oxoisoaporphine inhibitors for MAO-A and human parasite proteins. Eur J Med Chem 2011; 46: 5838-51.
Botella-Rocamora P, López-Quilez A, Martínez-Beneito MA. Spatial moving average risk smoothing. Stat Med 2013; 32: 2595- 612.
González-Díaz H, Duardo-Sánchez A, Ubeira FM, *et al.* Review of MARCH-INSIDE & Complex Networks Prediction of Drugs: ADMET, Anti-parasite Activity, Metabolizing Enzymes and Cardiotoxicity Proteome Biomarkers. Curr Drug Metab 2010; 11: 379-406.
González-Díaz H, Prado-Prado F, García-Mera X, *et al.* MIND- BEST: Web Server for Drugs and Target Discovery; Design, Synthesis, and Assay of MAO-B Inhibitors and Theoretical- Experimental Study of G3PDH Protein from Trichomonas gallinae. J Proteome Res 2011; 10: 1698-718.
Rodríguez-Soca Y, Munteanu CR, Dorado J, *et al.* Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions. J Proteome Res 20 I O; 9: 1182-90.
González-Díaz H, Romaris F, Duardo-Sánchez A, *et al.* Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, applications, and legal issues. Curr Pharrn Des 2010; 16: 2737-64.
González-Díaz H, Prado-Prado FJ, García-Mera X, *et al.* MIND- BEST: web server for drugs & target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretic-experimental study of G3PD protein from Trichomona gallineae. J Proteome Res 2011; 10(4): 1698-718.
Munteanu CR, Vázquez JM, Dorado J, *et al.* Complex network spectral moments for ATCUN motif DNA cleavage: first predictive study on proteins of human pathogen parasites. J Proteome Res 2009; 8: 5219-28.
Concu R, Dea-Ayuela MA, Pérez-Montoto LG, *et al.* Prediction of enzyme classes from 3D structure: a general model and examples of experimental-theoretic scoring of peptide mas s fingerprints of Leisbmania proteins. J Proteome Res 2009; 8: 4372-82.
Santana L, González-Díaz H, Quezada E, *et al.* Quantitative structure-activity relationship and complex network approach to monoamine oxidase a and B inhibitors. J Med Chem 2008; 51: 6740-51.
González-Díaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. Curr Top Med Chem 2008; 8: 1676-90.
Aguero-Chapin G, González-Díaz H, de la Riva G, *et al.* MMM- QSAR recognition of ribonucleases without alignment: comparison with an HMM model and isolation from Schizosaccharomyces pombe, prediction, and experimental assay of a new sequence. J Chem Inf Model 2008; 48: 434-48.
Aguero-Chapin G, Antunes A, Ubeira FM, *et al.* Comparative study of topological indices of macro/supramolecular RNA complex networks. J Chem Inf Model2008; 48: 2265-77.
González-Díaz H, Vilar S, Santana L, *et al.* Medicinal Chemistry and Bioinformatics - Current Trends in Drugs Discovery with Networks Topological Indices. Curr Top Med Chem 2007; 7: 1025- 39.
Ramos de Armas R, González-Díaz H, Molina R, *et al.* Markovian Backbone Negentropies: Molecular descriptors for protein research. I. Predicting protein stability in Are repressor mutants. Proteins 2004; 56: 715-23.
González-Díaz H, Marrero Y, Hernández I, *et al.* 3D-MEDNEs: an alternative "in silico" technique for chemical research in toxicology. l. prediction of chemically induced agranulocytosis. Chem Res Toxicol 2003; 16: 1318-27.
González-Díaz H, Duardo-Sánchez A, Ubeira FM, *et al.* Review of MARCH-INSIDE & complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. Curr Drug Metab 2010; 11: 379-406.

Batagelj V, Mrvar A. Pajek- Analysis and Visualization of Large Networks. Lect Notes Comput Sci 2002; 2265: 477-8.
González-Díaz H, González-Díaz Y, Santana L, *et al.* Proteomics, networks and connectivity indices. Proteomics 2008; 8: 750-78.
Santana L, Uriarte E, González-Díaz H, *et al.* A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. J Med Chem 2006; 49: 1149- 56.
González-Díaz H, Prado-Prado F, García-Mera X, *et al.* MIND- BEST: Web Server for Drugs and Target Discovery; Design, Synthesis, and Assay of MAO-B Inhibitors and Theoretical- Experimental Study ofG3PDH Protein from Trichomonas gallinae. J Proteome Res 2011; 10: 1698-718.
González-Díaz H, Saiz-Urra L, Molina R, *et al.* Computational chemistry approach to protein kinase recognition using 3D stochastic van der Waals spectral moments. J Comput Chem 2007; 28: 1042-8.
González-Díaz H, Molina R, Uriarte E. Recognition of stable protein mutants with 3D stochastic average electrostatic potentials, Febs Letters 2005; 579: 4297-301.
González-Díaz H, Pérez-Bello A, Uriarte E. Stochastic molecular descriptors for polymers. 3. Markov electrostatic moments as polymer 2D-folding descriptors: RNA--QSAR for mycobacterial promoters. Polymer 2005; 46: 6461-73.
Saiz-Urra L, González-Díaz H, Uriarte E. Proteins Markovian 3D- QSAR with spherically-truncated average electrostatic potentials, Bioorg Med Chem 2005; 13: 3641-47.
González-Díaz H, Uriarte E, Ramos de Armas R. Predicting stability of Are repressor mutants with protein stochastic moments, Bioorg Med Chem 2005; 13: 323-31.
Concu R, Podda G, Uriarte E, *et al.* Computational chemistry study of 3D-structure-function relationships for enzymes based on Markov models for protein electrostatic, HINT, and van der Waals potentials. J Comput Chem 2009; 30(9): 1510-20.
González-Díaz H, Saiz-Urra L, Molina R, *et al.* A Model for the Recognition of Protein Kinases Based on the Entropy of 3D van der Waals Interactions. J Proteome Res 2007; 6: 904-8.
González-Díaz H, Saiz-Urra L, Molina R, *et al.* Computational chemistry approach to protein kinase recognition using 3D stochastic van der Waals spectral moments. J Comput Chem 2007; 28: 1042-8.
González-Díaz H, Molina R, Uriarte E. Recognition of stable protein mutants with 3D stochastic average electrostatic potentials. FEBS Letters 2005; 579: 4297-301.
Concu R, Podda G, Uriarte E, *et al.* Computational chemistry study of 3D-structure-function retationships for enzymes based on Markov modets for protein etectrostatic, HINT, and van der Waals potentials. J Comput Chem 2009; 30: 1510-20.
González-Díaz H, Pérez-Castillo Y, Podda G, *et al.* Computational Chemistry Comparison of Stable/Nonstable Protein Mutants Classification Models Based on 3D and Topological Indices J Comput Chem 2007; 28: 1990-95.
Berca MN, Duardo-Sánchez A, González-Díaz H, *et al.* Markov entropy for biology, parasitology, linguistic, technology, social and law networks. In: González-Díaz H., Prado-Prado F. J., García-Mera X. eds). Complex Network Entropy: From Molecules to Biology, Parasitology, Technology, Social, Legal, and Neurosciences. Kerala, India: Transworld Research Network, 2011, 127-142.
Riera-Femández P, Munteanu CR, Pedreira-Souto N, *et al.* Definition of Markov-Harary Invariants and Review of Classic Topological Indices and Databases in Biology, Parasitology, Technology, and Social-Legal Networks. Curr Bioinfo 2011; 6: 94- 121.
Riera-Fernández I**,** Martin-Romalde R, Prado-Prado FJ, *et al.* From QSAR modets of Drugs to Complex Networks: State-of-Art Review and Introduction of New Markov-Spectral Moments Indices, Current Topics in Medicinal Chemistry 2012.
StatSoft.Jnc. STATISTICA (data analysis software system), version 6.0, www.statsoft.com.Statsoft, Inc., 2002.
Hopfietd LI. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 1982; 79: 2554-8.
Kier LB. Use of molecular negentropy to encode structure governing biological activity. J Pharm Sci 1980; 69: 807-10.
Agrawal VK, Khadikar PV. Modelling of carbonic anhydrase inhibitory activity of sulfonamides using molecular negentropy. Bioorg Med Chem Lett 2003; 13: 447-53.
Graham DJ, Schacht D. Base Information Content in Organic Molecular Formulae. J Chem Inf Comput Sci 2000; 40: 942.
Graham DJ. Information Content in Organic Molecules: Structure Considerations Based on Integer Statistics. J Chem Inf Comput Sci. 2002; 42: 215.
Graham DJ, Malarkey C, Schulmerich MV. Information Content in Organic Molecules: Quantification and Statistical Structure *via* Brownian Processing. J Chem Inf Comput Sci 2004; 44 (5): 1601- 11.
Graham DJ, Schulmerich MV. Information Content in Organic Molecules: Reaction Pathway Analysis *via* Brownian Processing, J Chem Inf Comput Sci 2004; 44 (5):1612-22.
Graham DJ. Information Content and Organic Molecules: Aggregation States and Solvent Effects, J Chem Inf Modet 2005; 45(5): 1223-36.

Graham DJ. Information Content in Organic Molecules: Brownian Processing at Low Levels. J Chem Inf Modet2007; 47: 376-89.
Stahura FL, Godden JW, Bajorath J. Differential Shannon entropy analysis identifies molecular property descriptors that predict aqueous solubility of synthetic compounds with high accuracy in binary QSAR calculations. J Chem Inf Comput Sci. 2002; 42: 550- 8.
Stahura FL, Godden JW, Xue L, *et al.* Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations. J Chem Inf Comput Sci. 2000; 40: 1245-52.
Roy K, Saha A. Comparative QSPR studies with molecular connectivity, molecular negentropy and TAU indices Part 1: Molecular thermochemical properties of diverse functional acyclic compounds. J Mol Mode12003; 9: 259-70.
Agrawal VK, Bano S, Khadikar PV. QSAR analysis of antibacterial activity of some 4-aminodiphenylsulfone derivatives. Acta Microbiol Immunol Hung 2003; 50: 385-93.
Agrawal VK, Karmarkar S, Khadikar PV. QSAR study on binding affinity of PATs (rodenticides) to the [3H]-mepyramine-labelled HI receptor in rat and guinea pig brain. SAR QSAR Environ Res 2001; 12: 529-45.
Agrawal VK, Khadikar PV. QSAR studies on acylated histamine derivatives. Bioorg Med Chem 2001; 9: 2787-92.
Katritzky AR, Lomaka A, Petrukhin R, *et al.* QSPR correlation of the metting point for pyridinium bromides, potential ionie liquids. J Chem Inf Comput Sci 2002; 42: 71-4.
Katritzky AR, Perumal S, Petrukhin R, *et al.* Codessa-based theoretical QSPR model for hydantoin HPLC-RT lipophilicities. 1 Chem Inf Comput Sci 2001; 41: 569-74.
Strait BJ, Dewey TG. The Shannon information entropy of protein sequences. Biopbys 1996; 71: 148-55.
Dima RI, Thirumalai D. Proteins associated with diseases sbow enhanced sequence correlation between charged residues. Bioinformatics 2004; 20: 2345-54.
Loewenstem D, Yianilos PN. Significantly lower entropy estimates for natural DNA sequences. J Comput Biol 1999; 6: 125-42.
Manke T, Demetrius L, Vingron M. Letbality and entropy of protein interaction networks. Genome Inform Ser 2005; 16: 159-63.
Mikolas P, Vyhnanek J, Skoch A, *et al.* Analysis of fMRI time- series by entropy measures. Neuro Endocrinol Lett 2012; 33: 471- 6.
Lee S, Park K, Kim D. Building a drug-target network and its applications. Expert Opin Drug Discov 2009; 4: 1-13.
Mestres I, Gregori-Puigjane E, Valverde S, *et al.* The topology of drug-target interaction networks: implicit dependence on drug properties and target families. Mol Biosyst 2009; 5: 1051-7.
Rosa da Silva M, Sun J, Ma HW, *et al.* Metabolic networks. In: Junker B. H., Schreiber F. eds). Analysis of biological networks. New Jersey: Wiley & Sons, 2008, 233-253.
Lee DS, Park J, Kay KA, *et al.* The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci U S A 2008; 105: 9880-5.
Huang T, Cben L, Cai YD, *et al.* Classification and analysis of regulatory pathways using graph property, biochemical and physicochemical property, and functional property. PLoS ONE 2011; 6: e25297.
Huang T, Sbi XH, Wang P, *et al.* Analysis and prediction of the metabolic stability of proteins based on their sequential features, subcellular locations and interaction networks. PLoS ONE 2010; 5: e10972.
Baldazzi V, Ropers D, Markowicz Y, *et al.* Tbe carbon assimilation network in Escherichia coli is densely connected and largely sign- determined by directions of metabolic fluxes. PLoS Comput Biol 2010; 6: e1000812.
Costa RS, Machado D, Rocba I, *et al.* Hybrid dynamic modeling of Escherichia coli central metabolic network combining Michaelis- Menten and approximate kinetic equations. Biosystems 2010; 100: 150-7.
Gerlee P, Lizana L, Sneppen K. Pathway identification by network pruning in the metabolic network of Escherichia coli. Bioinformatics 2009; 25: 3282-8.
Fowler ZL, Gikandi WW, Koffas MA. Increased malonyl coenzyme A biosynthesis by tuning the Escherichia coli metabolic network and its application to flavanone production. Appl Environ Microbiol2009; 75: 5831-39.
Konig R, Schramm G, Oswald M, *et al.* Discovering functional gene express ion patterns in the metabolic network of Escherichia coli with wavelets transforms. BMC Bioinformatics 2006; 7: 119.
Imielinski M, Betta C, Rubin H, *et al.* Systematic analysis of conservation relations in Escherichia coli genome-scale metabolic network reveals novel growth media. Biophys 1 2006; 90: 2659-72.
Lin H, Bennett GN, San KY. Chemostat culture characterization of Escherichia coli mutant strains metabolically engineered for aerobic succinate production: a study of the modified metabolic network based on metabolite profile, enzyme activity, and gene express ion profile. Metab Eng 2005; 7: 337-52.
Ghim CM, Gob KI, Kahng B. Lethality and synthetic lethality in the genome-wide metabolic network of Escherichia coli. J Theor Bio12005; 237: 401-11.
Schmid JW, Mauch K, Reuss M, *et al.* Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in Escherichia coli. Metab Eng 2004; 6: 364-77.

Ligbt S, Kraulis P. Network analysis of metabolic enzyme evolution in Escherichia coli. BMC Bioinformatics 2004; 5: 15.
Burgard AP, Maranas CD. Probing the performance limits of the Escherichia coli metabolic network subject to gene additions or deletions. Biotechnol Bioeng 2001; 74: 364-75.
Edwards JS, Palsson BO. Robustness analysis of the Escherichia coli metabolic network. Biotechnol Prog 2000; 16: 927-39.
Goffeau A. The yeast genome directory. Nature 1997; 387: 5.
Burglin TR, Lobos E, Blaxter ML. Caenorhabditis elegans as a model for parasitic nematodes. Int J Parasitol 1998; 28: 395-411.
Consortium TCeS. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 1998; 282: 2012-18.
Bird DM, Opperman CH. Caenorhabditis elegans: A Genetic Guide to Parasitic Nematode Biology. J Nematol1998; 30: 299-308.
Holden-Dye L, Walker RJ. Anthelmintic drugs. WormBook 2007: 1-13.
Muller B, Grossniklaus U. Model organisms-A historical perspective. 1 Proteomics 2010; 73: 2054-63.
Poulin R. Network analysis shining light on parasite ecology and diversity. Trends Parasitol 2010; 26: 492-8.
Kotter R. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. Neuroinforrnatics 2004; 2: 127-44.
Stephan KE, Kamper L, Bozkurt A, *et al.* Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). Philos Trans R Soc Lond B Biol Sci 2001; 356: 1159-86.
Mas-Coma S. Epidemiology of fascioliasis in human endemic areas. 1 Hetminthol2005; 9: 207-16.
González-Díaz H, Mezo M, González-Warleta M, *et al.* Network prediction of fasciolosis spreading in Galicia (NW Spain). In: González-Díaz H., Munteanu C. R. (eds). Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks. Kerala (India): Transworld Research Network, 2010,191-204.
Wasserman S, Faust K. Social network analysis: methods and applications. Cambridge: Cambridge University Press, 1999.
Fowler JH, Jeon S. The authority of Supreme Court precedent. Social Networks 2008; 30: 16-30.
Duardo-Sánchez A. Study of criminal law networks with Markov- probability centralities. In: González-Díaz H. (ed) Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks. Kerala, India: Bentham, 2010, 205-212.
Duardo-Sánchez A. Criminal law networks, markov chains, Shannon entropy and artificial neural networks. In: González-Díaz H. (ed) complex Network Entropy: From Molecules to Biology, Parasitology, Technology, Social, Legal, and Neurosciences. Kerala, India: Bentbam, 2011,107-114.
Yam KL, Takhistov PT, Miltz J. Intelligent Packaging: Concepts and Applications. J Food Sci 2005; 70: 1-10.
Suppakul P, Miltz J, Sonnevetd K, *et al.* Active Packaging Technologies with an Emphasis on Antimicrobial Packaging and its Applications. J Food Sci 2003; 68: 408-20.
Rucker G, Rucker C. Walk counts, labyrinthicity, and complexity of acyclic and cyclic grapbs and molecules. J Chem Inf Comput Sci 2000; 40: 99-106.
Rucker G, Rucker C. Substructure, subgraph, and walk counts as measures of the complexity of graphs and molecules. J Chem Inf Comput Sci 2001; 41: 1457-62.
Gutman I, Rucker C, Rucker G. On walks in molecular graphs. J Chem Inf Comput Sci 2001; 41: 739-45.
Lukovits I, Trinajstic N. Atomic walk counts of negative order. J Cbem Inf Comput Sci 2003; 43: 1110-14.
González-Díaz H, Riera-Fernández P, Pazos A, *et al.* The Rucker- Markov invariants of complex Bio-Systems: Applications in Parasitology and Neuroinformatics, Biosystems 2013.
Moreau G, Broto P. Autocorretation of a topological structure: A new molecular descriptor. Nouv J Chim 1980; 4: 359-60.
Moreau G, Broto P. Autocorretation of molecular structures, application to SAR studies. Nouv J Chim 1980; 4: 757-64.
Broto P, Moreau G, Vandycke C. Molecular structures: perception, autocorrelation descriptor and sar studies: system of atomic contributions for the calculation of the n-octanol/water partition coefficients. Euro J Med Chem 1984; 19: 71-8.
Moro S, Bacilieri M, Cacciari B, *et al.* Autocorretation of molecular electrostatic potential surface properties combined with partial least squares analysis as new strategy for the prediction of the activity of human A(3) adenosine receptor antagonists. J Med Chem 2005; 48: 5698-704.
Fernández L, Caballero J, Abreu JL, *et al.* Amino Acid Sequence Autocorrelation Vectors and Bayesian-Regularized Genetic Neural Networks for Modeling Protein Conformational Stability: Gene V Protein Mutants. Proteins 2007; 67: 834-52.
Caballero J, Garriga M, Fernández M. 2D Autocorrelation modeling of the negative inotropic activity of calcium entry blockers using Bayesian-regularized genetic neural networks. Bioorg Med Chem 2006; 14: 3330-40.

Caballero J, Fernández L, Garriga M, *et al.* Proteometric study of ghretin receptor function variations upon mutations using amino acid sequence autocorrelation vectors and genetic algorithm-based least square support vector machines. J Mol Graph Model 2007; 26: 166-78.
Caballero J, Fernández L, Abreu JL, *et al.* Amino Acid Sequence Autocorrelation vectors and ensembles of Bayesian-Regularized Genetic Neural Networks for prediction of conformational stability of human lysozyme mutants. J Chem InfModet2006; 46: 1255-68.
Han P, Zhang X, Norton RS, *et al.* Large-scale prediction of long disordered regions in proteins using random forests. BMC Bioinformatics 2009; 10: 8.
Li ZR, Lin HH, Han L Y, *et al.* PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. Nucleic Acids Res 2006; 34: W32-37.
Munteanu CR, González-Díaz H, Borges F, *et al.* Natural/random protein classification models based on star network topological indices. J Theor Biol 2008; 254: 775-83.
Modha OS, Singh R. Network architecture of the long-distance pathways in the macaque brain. Proc Natl Acad Sci U S A 2010; 107: 13485-90.
Wiener H. Structural determination of paraffin boiling points. J Am Chem Soc 1947; 69: 17-20.
Wiener H. Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons. J Am Chem Soc 1947; 69: 2636-8.
Wiener H. Relation of the physical properties of the isomeric alkanes to molecular structure. Surface tension, specific dispersion, and critical solution temperature in aniline. J Phys Colloid Chem 1948; 52: 1082-9.
Wiener H. Vapor pressure-temperature relationships among the branched paraffin hydrocarbons. J Phys Colloid Chem 1948; 52: 425-30.
Hosoya H. Topological index, a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. Bull Chem Soc Jpn 1971; 44: 2332-9.
Hosoya H. Mathematical meaning and importance of the topological index Z. Croat Chem Acta 2007; 80: 239-49.
Consonni V, Todeschini R. Molecular descriptors. In: Puzyn *T.*,Leszczynski J., Cronin M. T. O. (eds). Recent advances in QSAR studies: Methods and applications. Springer, 2010, 29-102.
Harary F. Status and contrastatus. Sociometry 1959; 22: 23-43.
Diudea MV, Gutman I. Wiener-type topological indices, Croatica Chemica Acta 1998;71 :21-51.
Rosenblatt F. Principles of neurodynamics; perceptrons and the theory of brain mechanisms. Washington: Spartan Books, 1962.
TBA. Highly Discriminating Distance-Based topological index. Chem Phys Lett 1981; 89: 399-404.
Sharma SK, Kumar P, Narasimhan B, *et al.* Synthesis, antimicrobial, anticancer evaluation and QSAR studies of. Eur J Med Chem 2012; 48: 16-25.
Thakur A, Thakur M, Khadikar PV, *et al.* QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: topological. Bioorg Med Chem 2004; 12: 789-93.
Krawczuk A, Voetket A, Lulek J, *et al.* Use of topological indices of polychlorinated biphenyls in structure-retention. J Chromatogr A 2003; 1018: 63-71.
Yadav S, Kumar P, De Clercq E, *et al.* 4-[I-(Substituted aryl/alkyl carbonyl)-benzoimidazol-2-yl]-benzenesulfonic acids. Eur J Med Chem 2010; 45: 5985-97.
Naik PK, Dubey A, Kumar R. Development of predictive quantitative structure-activity relationship models of. J Biomol Screen 2010; 15: 1194-203.
Fernandes JP, Pasqualoto KF, Fetli VM, *et al.* QSAR modeling of a set of pyrazinoate esters as antituberculosis prodrugs. Arch Pharm 2010; 343: 91-7.
Panaye A, Doucet JP, Devillers J, et *al.* Decision trees *versus* support vector machine for classification of androgen. SAR QSAR Environ Res 2008; 19: 129-51.
Ma XL, Chen C, Yang J. Predictive model of blood-brain barrier penetration of organic compounds. Acta Pharmacol Sin 2005; 26: 500-12.
Dashtbozorgi Z, Golmoharnmadi H. Quantitative structure-property relationship modeling of water-to-wet butyl. J Sep Sci 2010; 33: 3800-10.
Jalali-Heravi M, Fatemi MH. Prediction of thermal conductivity detection response factors using an artificial neural network. J Chromatogr A 2000; 897: 227-35.
Dehrner M, Grabner M, Varmuza K. Information indices with high discriminative power for graphs. PLoS One 2012; 7: e31214.
Basak SC, Milis O, Gute BD, *et al.* Use of mathematical structural invariants analyzing combinatorial libraries: a. Curr Comput Aided Drug Des 2010; 6: 240-51.
Ratkiewicz A, Truong TN. Application of chemical graph theory for automated mechanism generation. J Chem Inf Comput Sci 2003; 43: 36-44.
Randic M, Pompe M. The variable molecular descriptors based on distance related matrices. J Chem Inf Comput Sci 2001; 41: 575- 81.
Duardo-Sánchez A, Munteanu CR, Pazos A, *et al.* ANN Modeting of Complex Networks of Biochemical Reactions, Ecosystems, and U.S. Supreme Court Citations with New Balaban-Markov Centralities, Complexity 2013:submitted.
Fowler JH, Jeon S. The authority of Supreme Court precedent. Social Networks 2008; 30: 16-30.