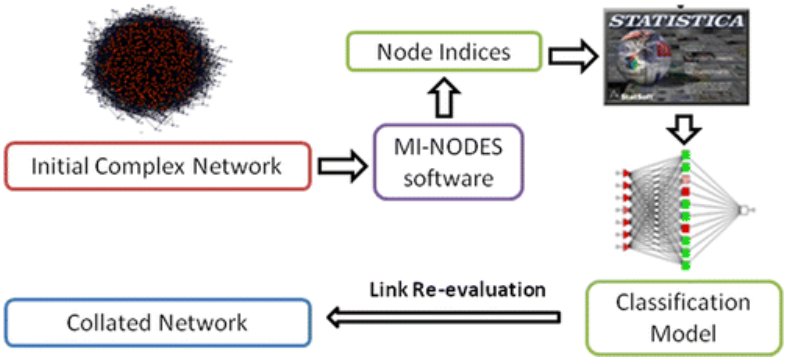
Journal of Chemical Information and Modeling. 2014; 54(1): 16-29

Modeling complex metabolic reactions, ecological systems, and financial and legal networks with MIANN models based on Markov-Wiener node descriptors

Aliuska Duardo-Sánchez†‡, Cristian R. Munteanu*†, Pablo Riera-Fernández†, Antonio López-Díaz‡, Alejandro Pazos†, and Humberto González-Díaz§‖

† *Department of Information and Communication Technologies, Computer Science Faculty, University of A Coruña, Campus de Elviña, 15071, A Coruña, A Coruña, Spain*
‡ *Department of Special Public Law, Financial and Tributary Law Area, Faculty of Law, University of Santiago de Compostela (USC), 15782, Santiago de Compostela, A Coruña, Spain*
§ *Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), 48940, Leioa, Bizkaia, Spain*
‖ *IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Biscay, Spain*

Abstract



The use of numerical parameters in Complex Network analysis is expanding to new fields of application. At a molecular level, we can use them to describe the molecular structure of chemical entities, protein interactions, or metabolic networks. However, the applications are not restricted to the world of molecules and can be extended to the study of macroscopic nonliving systems, organisms, or even legal or social networks. On the other hand, the development of the field of Artificial Intelligence has led to the formulation of computational algorithms whose design is based on the structure and functioning of networks of biological neurons. These algorithms, called Artificial Neural Networks (ANNs), can be useful for the study of complex networks, since the numerical parameters that encode information of the network (for example centralities/node descriptors) can be used as inputs for the ANNs. The Wiener index (W) is a graph invariant widely used in chemoinformatics to quantify the molecular structure of drugs and to study complex networks. In this work, we explore for the first time the possibility of using Markov chains to calculate analogues of node distance numbers/W to describe complex networks from the point of view of their nodes. These parameters are called Markov-Wiener node descriptors of order k^{th} (W_k). Please, note that these descriptors are not related to Markov-Wiener stochastic processes. Here, we calculated the $W_k(i)$ values for a very high number of nodes (>100,000) in more than 100 different complex networks using the software MI-NODES. These networks were grouped according to the field of application. Molecular networks include the Metabolic Reaction Networks (MRNs) of 40 different organisms. In addition, we analyzed other biological and legal and social networks. These include the Interaction Web Database

Biological Networks (IWDBNs), with 75 food webs or ecological systems and the Spanish Financial Law Network (SFLN). The calculated $W_k(i)$ values were used as inputs for different ANNs in order to discriminate correct node connectivity patterns from incorrect random patterns. The MIANN models obtained present good values of Sensitivity/Specificity (%): MRNs (78/78), IWDBNs (90/88), and SFLN (86/84). These preliminary results are very promising from the point of view of a first exploratory study and suggest that the use of these models could be extended to the high-throughput re-evaluation of connectivity in known complex networks (collation).

1 Introduction

1.1The Classic Wiener Index

In the last part of the nineteenth century and in the twentieth century the interest for the study of the molecular structure led to the formulation of questions about how to encode and quantify the information contained in the molecule. As a result of these questions, the concept of molecular descriptor was defined as "the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment".(1) In 1947, Wiener published an article entitled *Structural determination of paraffin boiling points*.(2) 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. Therefore, it is possible to find an equation that relates the structure of the studied paraffins with their boiling points. This equation, also used by Wiener in refs 3−5, can be written in a general form as

$$\Delta t = \frac{a}{n^2} \cdot \Delta w + b \cdot \Delta p \tag{1}$$

where n is the number of atoms, p is the polarity number, defined as the number of pairs of carbon atoms which are separated by three carbon bonds, and w is the path number, defined as the sum of the distances between any two carbon atoms in the molecule and considered as one of the oldest topological indices. This last term was coined by Hosoya in 1971 to refer to the Z index,(-6, 7) and it is currently used to define all the numerical quantifiers of molecular topology that are mathematically derived from the structural graph of a molecule, usually an H-depleted molecular graph.(8) The path number is also called Wiener index or Wiener number (W), and it is calculated as the half sum of all the elements d_{ij} of the distance matrix (D). As it can be seen, more distant atom pairs make a larger contribution to W than adjacent atom pairs:

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

It is interesting to point out that the Wiener index was independently proposed in 1959 by Harary in the context of sociometry, with the name *total status of a graph*(9) as well as in 1975 by Rouvray and Crafford.(10) This fact indicates that the Wiener index concept was not well-known in those times. However, in the middle of the 1970s many authors began to study the properties and applications of molecular descriptors. This led to the development of new topological indices (TIs), some of which were based on W. A complete list and a brief explanation of these indices can be found in ref 11. Taking into consideration the classification proposed by Balaban for the TIs,(12) W (together with Z) belongs to the first of three generations or classes. First-generation TIs are integer numbers based on integer local graph-vertex invariants (LOVI) and have a high degeneracy that limits their use.

According to Diudea and Gutman,(10) the physical and chemical properties of organic substances, which can be expected to depend on the area of the molecular surface and/or on the branching of the molecular carbon-atom skeleton, are usually well correlated with *W*. Among them, there are the heats of formation, vaporization and atomization, density, boiling point, critical pressure, refractive index, surface tension and viscosity of various, acyclic and cyclic, saturated and unsaturated as well as aromatic hydrocarbon species, velocity of ultrasound in alkanes and alcohols, rate of electroreduction of chlorobenzenes, etc. Correlations between *W* and melting points were also reported, but in this case the results were not completely satisfactory. *W* was also used to predict the behavior of organic substances in gas chromatography, for instance, chromatographic retention times (CRT) of monoalkyl- and *o*-dialkylbenzenes. In this sense, *W* is very useful in chemoinformatics for the search of models that connect the molecular structure with molecular properties.

1.2Applications of the Wiener Index in Chemoinformatics

In pharmaceutical design, we can find many applications of the Wiener index. For instance, Mandloi et al.(13) investigated the correlation of Wiener (*W*), Szeged (Sz), and molecular connectivity indices ($^{0}\chi_R$, $^{1}\chi_R$, and $^{2}\chi_R$) with molecular properties. Log P values of benzoic acid and its nuclear-substituted derivatives were used for this purpose. The statistical analyses for univariate and multivariate correlations indicated that both *W* and Sz are closely related to the connectivity indices ($^{m}\chi_R$) and that *W*, Sz, and $^{1}\chi_R$ have similar modeling potentials ($^{1}\chi_R$ gives slightly better results than both *W* and Sz). $^{0}\chi_R$ and $^{2}\chi_R$ are poorly correlated with log P. Lukovits established correlations between *W* and cytostatic and antihistaminic activities of certain pharmacologically interesting compounds as well as between *W* and their Estron-binding affinities.(14) He also employed *W* in the study of the *n*-octanol/water partition coefficient.(15)

Mendiratta and Madan(16) studied the relationship between *W* and the antiviral activity of a series of 118 5-vinylpyrimidine nucleoside analogues. The predicted activity of each compound was compared with reported antiviral activity against herpes simplex virus type I. Due to the significant correlation between antiviral activity and *W*, it was possible to predict antiviral activity with an accuracy of 83%.

In the work carried out by Agrawal et al.(17) the antimalarial activity of a series of sulfonamide derivatives (2,4-diamino-6-quinazoline sulfonamides) was modeled topologically using *W* and Sz. It was observed that the models based on *W* gave slightly better results than the models based on Sz. Sardana and Madan(18) studied the relationship of the molecular connectivity index ($^{1}\chi$), *W*, and the eccentric connectivity index (ξ^c) with the diuretic activity of 68 sulfamoylbenzoic acid derivatives. The models had an 82% accuracy rate in $^{1}\chi$, an 85% accuracy rate in *W*, and a 90% accuracy rate in ξ^c. In another work, the relationship of *W*, Zagreb group parameter (M1), and ξ^c with the anticonvulsant activity of a series of 41 substituted benazamides/benzylamines was investigated.(19) The models had an 88% (M1), 94% (ξ^c), and 97% (*W*) accuracy rate.

Gupta et al. studied(20) the relationship of ξ^c and *W* with regard to anti-inflammatory activity for a data set consisting of 76 pyrazole carboxylic acid hydrazide analogues. A prediction with a 90% accuracy rate was obtained using ξ^c and an 84% accuracy rate in the case of *W*. Bajaj et al.(21) studied the relationship of the Wiener topochemical index (a modification of *W* sensitive to the presence of heteroatoms and with less degeneracy) and Wiener index with the anti-HIV activity of 62 phenethylthiazolethiourea compounds. The prediction accuracy rate was 90% in both cases. The relationship of anti-HIV activity of 61 acylthiocarbamates with *W*, $^{1}\chi$, and ξ^c was also investigated by these authors.(22) 95% (ξ^c), 97% ($^{1}\chi$), and 98% (*W*) accuracy rates were observed. In another work, the relationship of anti-inflammatory activity of 112 N-arylanthranilic acids with *W*, Zagreb indices M1 and M2, and ξ^c was studied.(23) The different models had an 82.6% (ξ^c), 86.8% (*W*), 88.88% (M1), and 90.3% (M2) accuracy rate

In the area of cancer research, the inhibition of CDK2/cyclin A by 42 3-aminopyrazoles was studied using *W*, the atomic molecular connectivity index (χ^A), and the superadjacency topochemical index ($\int^{Ac}$).(24) The different models had an 86% (*W*), 88% ($\int^{Ac}$), and 89% (χ^A) accuracy rate. With the aim to develop methods to select drug candidates for the treatment of Alzheimer's disease, Kumar and Madan studied(25) the relationship of *W*, M1, and ξ^c with the glycogen synthase kinase-3 beta inhibitory activity of 28 thiadiazolidinones. The prediction accuracy rate was 83% (M1), 86% (ξ^c),

and 87% (*W*). Finally, Lather and Madan studied(26) the relationship between *W* and multidrug-resistance-associated protein inhibitory activity of 82 pyrrolopyrimidines and their derivatives. The prediction accuracy rate of the model was 88%. As we can see in these examples, the Wiener index has a wide range of applications in predictive studies, and, since it is one of the first topological indices, it is used in many works in order to compare the performance of new introduced indices.

1.3Complex Networks and MARCH-INSIDE Models

Graph and Complex Network theory is expanding its application to different levels of matter organization such as molecular, biological, technological, and social networks.(27-29) A network is a set of items, usually called *nodes*, with connections between them, which are called *links* or *edges*.(30) The nodes can be atoms, molecules, proteins, nucleic acids, drugs, cells, organisms, parasites, people, words, laws, computers, or any other part of a real system. The edges or links are relationships between the nodes, such as chemical bonds, physical interactions, metabolic pathways, pharmacological actions, law recurrence, or social ties.(31-39)

On the other hand, there are many different experimental and/or theoretical methods to assign node–node links depending on the type of network we want to construct. Unfortunately, many of these methods are expensive in terms of time or resources. In addition, different methods that link nodes in the same type of network are not totally accurate and consequently they do not always coincide. A possible solution to this problem is the use of node descriptors of known networks as inputs of predictive models.(40) The reasons for using re-evaluations of link connectivity in networks are the following:

1. The experimental networks can have errors due to experiment conditions, calibrations, human errors, etc.
2. There are networks where the connectivity is just a prediction or it is the result of text data mining techniques (all involve possible errors).
3. The model that can re-evaluate the node connectivity can be used for new nodes as an alternative to the expensive and time-consuming experiments. In some cases, such as the interaction of all the possible pairs – triples of molecules, it is impossible to be carried out experimentally.
4. Contradictory information for nodes and links for different networks.

In fact, the use of predictive models in which the inputs are graph parameters is not limited to the study of molecules and has been extended to other complex systems.(41, 42) The first and one of the most studied TIs is the Wiener index, and it is possible to use Markov Chains (MC) to calculate it locally or globally within a graph considering all possible branches at different topological distances. The information is quantified in terms of $W_k(j)$ values, which are called Markov-Wiener node descriptors of order k^{th} for all j^{th} states (nodes) of an MC associated with the system. This MC is expressed by a Markov or Stochastic matrix ($\mathbf{\Pi}_1$) and represented by a graph of the studied system. The elements of $\mathbf{\Pi}_1$ are the probabilities ${}^1p_{ij}$ with which the i^{th} and j^{th} nodes connect to each other (there is a physical or functional tie, link, or relationship) within a graph. By using Chapman-Kolmogorov equations it is straightforward to realize the way to calculate $W_k(j)$ values for all nodes in a graph. We can use these values directly or sum some of them to obtain total or local parameters. Our group has introduced the software called MARCH-INSIDE (Markovian Chemicals In Silico Design), which has become a very useful tool for predictive studies on drugs, proteins, and more complex systems.(43-57) This software can calculate 1D (sequence), 2D (connectivity in the plane), and 3D (connectivity in the space) MC parameters, including $W_k(j)$ values, for many molecular systems. MARCH-INSIDE is able to characterize small molecules (drugs, metabolites, organic compounds), biopolymers (gene sequence, protein sequence or 3D structure, and RNA secondary structure) and artificial polymers, but it can perform a limited manage of other complex networks. This occurs because MARCH-INSIDE can read, transform into Markov matrix, represent as graph, and calculate the Wiener index for molecular formats (.mol or SMILE .txt files for drugs, .pdb for proteins, or .ct files for RNAs), but it is unable to upload formats of Complex Networks (.mat, .net, .dat, .gml, etc.).

Consequently, we have reprogrammed the MARCH-INSIDE application, creating new software able to manage complex networks. The new program is called MI-NODES (MARCH-INSIDE NOde DEScriptors), and it is compatible with other programs like Pajek or CentiBin, since it is able to read .mat, .net, and .dat formats. A very interesting feature of MI-NODES is that it can process multiple networks and calculate both MC global TIs and/or node descriptors for all these networks. It is also able to export them in a single file in network-by-network and/or node vs node output formats. The classic TIs can include additional information such as Markov node linkage probability (p_{ij}) for any i, j nodes of a graph. In previous works, we have introduced other types of Markov TIs (and the respective node descriptors): Markov-Shannon Entropy node descriptors,(58) Markov-Randić indices,(59) Markov-Rücker indices,(60) Markov-Galvez indices,(61) Markov-Autocorrelation node descriptors,(62) and Markov-Harary numbers.(63) In these previous studies, we have used the indices, calculated with MI-NODES, in order to compare several types of complex networks from different fields such as biology, linguistics, technology, sociology, and law.

1.4MIANN Models

The methods used to predict structure–property relationships in complex systems (molecular or not) can be classified into two types: methods of type (1), used to quantify the structure of the system and methods of type (2), able to link the structure of the system with a property of this system (and others). Several methods of type (1) use Quantum Mechanics (in Molecular Sciences) and/or Graph theory (in Molecular and Social Sciences as well), whereas the methods of type (2) use Statistical and/or Machine Learning (ML) techniques.(64-68) Many computer programs implement type (1) and/or type (2) methods with applications in Molecular Sciences and/or a wide range of areas depending on the flexibility of the algorithms used. For instance, DRAGON,(69-71) TOPS-MODE,(72-75) TOMOCOMD,(76, 77) CODESSA,(78, 79) and MOE(80) are classic programs used to apply type (1) methods in Molecular Sciences. CentiBin,(81) Pajek,(82) or MI-NODES implement type (1) methods with applications in almost all areas of sciences but at the cost of simplification of detailed representation of the system. On the other hand, the Linear Discriminant Analysis (LDA) implemented in STATISTICA(83) or the ML methods implemented in WEKA(84) are examples of type (2) methods with widespread applications. In this context, different researchers/journals have edited important monographic issues in order to discuss different computational methods. For instance, Bisson has edited a special issue about Computational Chemogenomics in drug design and discovery.(85) Speck-Planche and Cordeiro guest-edited a special issue about computer-aided, synthesis and assay of anticancer agents.(86) Prado-Prado and García-Mera have also guest-edited a special issue about computer-aided drug design and molecular docking for disorders of the central nervous system and other diseases.(87) González-Díaz has guest-edited two special issues about multitarget models and Complex Networks applied to medicinal chemistry.(88, 89) In all these issues, and others of the same journal, several review and research papers in this area(52, 80, 90-123) have been published.

In particular, the bioinspired Artificial-Intelligence (AI) algorithms called Artificial Neural Networks (ANNs) are among the most powerful type (2) methods. As we mentioned in the previous section, MARCH-INSIDE (MI from now on) is a well-known type (1) method mentioned in many recent works published by different groups.(39, 52, 124-130) We can combine MI with different Machine Learning algorithms. In particular, we can combine MI with ANNs in order to seek predictive models. The name of this strategy is MIANN (MI and ANN models). In a recent paper, we have reviewed the MIANN strategy including theoretical basis, implementation in Web servers, and examples of applications in molecular sciences.(131) We have also developed new MIANN models for drug-target interactions, several physicochemical properties of surfactants, and large reaction networks in organic synthesis.

In the present work we introduce for the first time a new type of Wiener-like indices called Wiener-Markov node descriptors. This algorithm of type (1) is implemented in MI-NODES. Then, we use for the first time the MIANN strategy to study complex biomolecular, ecological, and social and legal systems using the Wiener-Markov node descriptors as input. In order to illustrate the use of the new method we have carried out three studies. In each study, we report for the first time a new model useful to re-evaluate the connectivity quality of different types of networks. Although very different systems were studied, the same workflow was used in all the experiments (see Figure 1**)**.

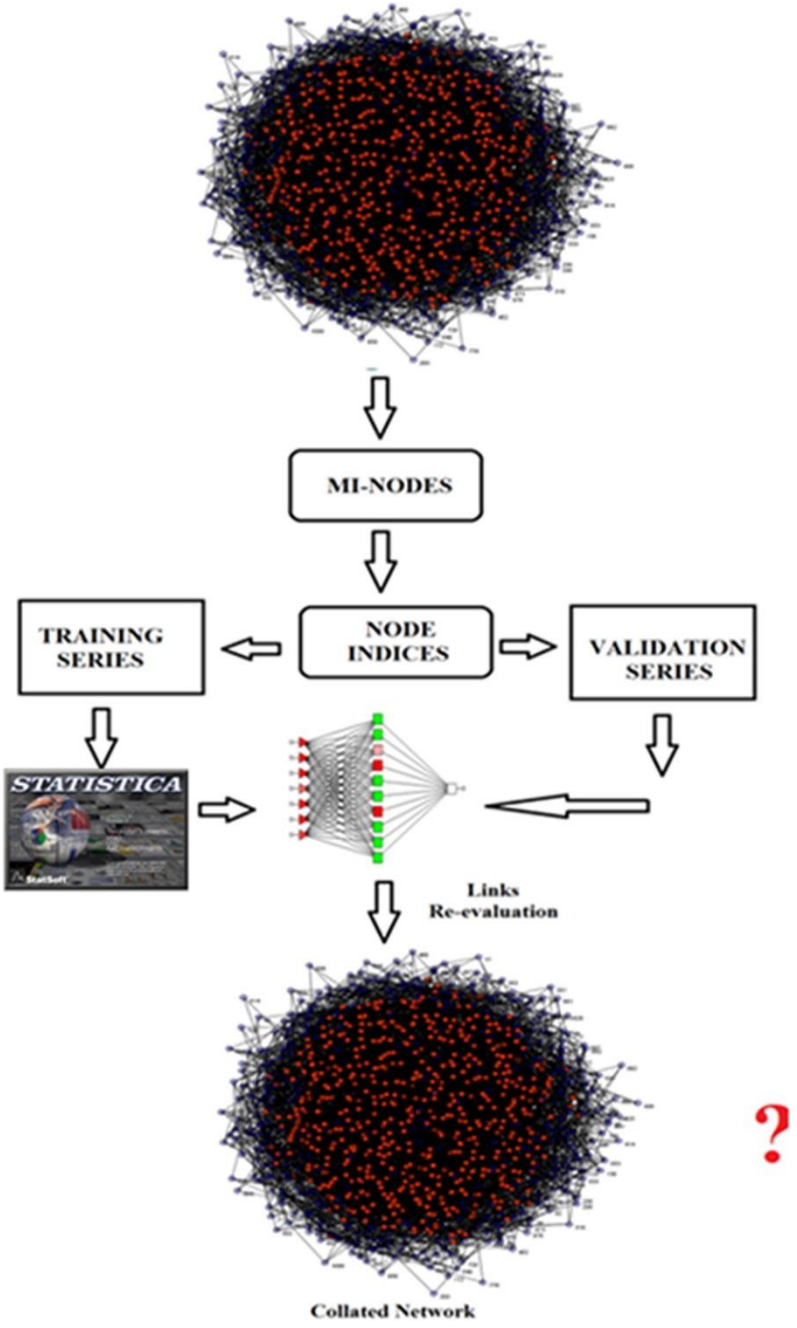


Figure 1. MIANN workflow example: blue/red nodes are training/validation cases (dark/light in gray scale).

The idea is to search for a MIANN model that uses the $W_k(j)$ values calculated for the nodes of a complex network as inputs of ANNs to decide which nodes are correctly linked. This class of model will allow us to computationally re-evaluate all the links of nodes in any complex network so that we do not have to rely upon experimentation to confirm the existence or not of a link between all pairs of links. By using this model, we should experimentally confirm only those connections predicted by the model with low link score and/or simply remove them from the network depending on the cost/benefit ratio. This work is proposing three studies: each study is proposing a prediction model based on several networks of the same network type. In the first study, we processed the full set of metabolic reactions of different organisms (bacteria, yeast, nematode, and plants). The node descriptors from 40 networks represented the model data set. In the second study, we used different biological and ecological networks, including predator–prey, parasite-host, plant-seed disperser,

anemone-clown fish species, and others. In the last study, we illustrate the application of the method to a complex network that takes into account all the historical record (1940–2004) of the Spanish Financial Law system (legal and social network). With the advent of the age of complex system sciences, this work can be considered as a basis for a relatively little studied but very important field: the assessment of the connectivity quality in new complex networks.

2 Materials and Methods

2.1Markov-Wiener Node Descriptors

The classic Markov matrix ($^1\mathbf{\Pi}$) for each network is constructed as follows: first, we download from the Internet the connectivity matrix **L** or the data about the links between the nodes to assemble **L** ($n \times n$ matrix, where n is the number of vertices). Next, the Markov matrix **Π** is built. It contains the vertices probability (p_{ij}) based on **L**. The probability matrix is raised to the power k, resulting $(^1\mathbf{\Pi})^k$. The resulting matrices $^k\mathbf{\Pi}$ are the k^{th} natural powers of $^1\mathbf{\Pi}$ and 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 W to general Markov-Wiener indices of order k^{th} is straightforward to carry out by multiplying the values of d_{ij} (distances obtained from the distance matrix D) by these probabilities $^kp_{ij}$. Therefore, we can obtain k values of the new Markov-Wiener indices $W_k(G)$ for a graph G, instead of only one Wiener index value obtained with the classic formulation. In addition, we can run the sum only over all the j^{th} nodes linked to one specific node i (the number of these nodes is symbolized here as $j \rightarrow i$ and it is equal to δ_i, the degree of i). In this simple case we can obtain a total of k values of new Markov-Wiener node descriptors, $W_k(i)$, for the node i^{th}:

$$W_k(G) = \frac{1}{2} \cdot \sum_{i=1}^{\mathbf{\Pi}\mathrm{D}} \sum_{J=1}^{\mathbf{\Pi}\mathrm{D}} {}^kp_{ij} \cdot d_{ij} \tag{3}$$

$$W_k(i) = \frac{1}{2} \cdot \sum_{i=1}^{1} \sum_{J=1}^{\boldsymbol{\delta_i}} {}^kp_{ij} \cdot d_{ij} = \frac{1}{2} \cdot \sum_{j \rightarrow i}^{\delta_i} {}^kp_{ij} \cdot d_{ij} \tag{4}$$

2.2Data sets Used

2.2.1Metabolic Reaction Networks (MRNs)

The data were downloaded directly from Barabasi's group Web site (http://www.nd.edu/~networks/resources.htm) as gzipped ASCII files. In these files each number represents a substrate in the metabolic network. Data-format is as follows: From → To (directed link). The information studied was previously obtained by Jeong et al. from the 'intermediate metabolism and bioenergetics' portions of the WIT database and used in order to try to understand the large-scale organization of metabolic networks.(132) According to the authors, the biochemical reactions described within the WIT database are composed of substrates and enzymes connected by directed links. For each reaction, educts and products were considered as nodes connected to the temporary educt–educt complexes and associated enzymes. Bidirectional reactions were considered separately. For a given organism with N substrates, E enzymes, and R intermediate complexes the full stoichiometric interactions were compiled into an (N+E+R) X (N+E+R) matrix, generated separately for each of the different organisms. Table 1 shows a summary of the properties of the MRNs studied. The names, abbreviations, and links for all the networks studied are as follows: *Aquifex aeolicus* = Aae; *Actinobacillus actinomycetemcomitans* = Aac; *Archaeoglobus fulgidus* = Afu; *Aeropyrum pernix* = Ape; *Arabidopsis thaliana* = Ath; *Borrelia burgdorferi* = Bbu; *Bacillus subtilis* = Bsu; *Clostridium acetobutylicum* = Cac; *Caenorhabditis elegans* = Cel; *Campylobacter*

jejuni = Cje; *Chlorobium tepidum* = Cte; *Chlamydia pneumoniae* = Cpn; *Chlamydia trachomatis* = Ctr; *Synechocystis sp.* = Csp; *Deinococcus radiodurans* = Dra; *Escherichia coli* = Eco; *Enterococcus faecalis* = Efa; *Emericella nidulans* = Eni; *Haemophilus influenza* = Hin; *Helicobacter pylori* = Hpy; *Mycobacterium bovis* = Mbo; *Mycoplasma genitalium* = Mge; *Methanococcus jannaschii* = Mja; *Mycobacterium leprae* = Mle; *Mycoplasma pneumonia* = Mpn; *Mycobacterium tuberculosis* = Mtu; *Neisseria gonorrheae* = Ngo; *Neisseria meningitidis* = Nme; *Oryza sativa* = Osa; *Pseudomonas aeruginosa* = Pae; *Pyrococcus furiosus* = Pfu; *Porphyromonas gingivalis* = Pgi; *Pyrococcus horikoshii* = Pho; *Streptococcus pneumonia* = Spn; *Rhodobacter capsulatus* = Rca; *Rickettsia prowazekii* = Rpr; *Saccharomyces cerevisiae* = Sce; *Streptococcus pyogenes* = Spy; *Methanobacterium thermoautotrophicum* = Mth; *Thermotoga marítima* = Tma; *Treponema pallidum* = Tpa; *Salmonella typhi* = Sty; *Yersinia pestis* = Ype.

Table 1. Average Values $W_k(i)_{org.avg}$ of Metabolic Networks of 43 Organisms vs Classic Parameters[a]

organism	N	L_{in}	L_{out}	R	E	g_{in}	g_{out}	D	W_1	W_2	W_3	W_4	W_5
Aae	419	1278	1249	401	285	2.1	2.2	3.3	0.87	1.08	1.26	1.44	1.57
Aac	395	1202	1166	380	271	2.1	2.2	3.2	0.88	1.07	1.24	1.43	1.56
Afu	496	1527	1484	486	299	2.2	2.2	3.5	0.85	1.09	1.29	1.48	1.61
Ape	204	588	575	178	135	2.2	2.2	3.2	0.95	1.11	1.25	1.46	1.6
Ath	302	804	789	250	185	2.1	2.3	3.5	0.89	1.12	1.3	1.48	1.62
Bbu	187	442	438	140	106	2.3	2.4	3	0.8	0.99	1.18	1.37	1.49
Bsu	785	2794	2741	916	516	2.2	2.1	3.3	0.8	1.09	1.3	1.52	1.65
Cac	494	1624	1578	511	344	2.1	2.2	3.3	0.83	1.08	1.28	1.46	1.59
Cel	462	1446	1418	450	295	2.1	2.2	3.3	0.9	1.12	1.32	1.51	1.65
Cje	380	1142	1115	359	254	2.1	2.3	3.2	0.88	1.09	1.27	1.45	1.58
Cte	389	1097	1062	333	231	2.1	2.2	3.3	0.88	1.1	1.3	1.51	1.63
Cpn	194	401	391	134	84	2.2	2.3	3.4	0.99	1.14	1.27	1.47	1.62
Ctr	215	479	462	158	94	2.2	2.4	3.5	0.9	1.06	1.22	1.38	1.5
Csp	546	1782	1746	570	370	2	2.2	3.3	0.88	1.13	1.33	1.56	1.68
Dra	815	2870	2811	965	557	2.2	2.1	3.3	0.89	1.12	1.31	1.52	1.65
Eco	778	2904	2859	968	570	2.2	2.1	3.2	0.79	1.03	1.24	1.44	1.57
Efa	386	1244	1218	382	281	2.1	2.2	3.1	0.81	1.04	1.24	1.42	1.55
Eni	383	1095	1081	339	254	2.1	2.2	3.3	0.89	1.11	1.31	1.5	1.65
Hin	526	1773	1746	597	361	2.1	2.3	3.2	0.77	1.05	1.26	1.48	1.59
Hpy	375	1181	1144	375	246	2	2.3	3.3	0.89	1.11	1.3	1.5	1.62
Mbo	429	1247	1221	391	282	2.2	2.2	3.2	0.87	1.09	1.27	1.46	1.6
Mge	209	535	525	196	85	2.4	2.2	3.5	0.96	1.14	1.26	1.38	1.48
Mja	424	1317	1272	415	264	2.2	2.3	3.5	0.88	1.11	1.29	1.47	1.6
Mle	422	1271	1244	402	282	2.2	2.2	3.2	0.83	1.06	1.25	1.44	1.58
Mpn	178	470	466	154	88	2.3	2.2	3.2	0.91	1.11	1.29	1.46	1.59
Mtu	587	1862	1823	589	358	2	2.2	3.3	0.88	1.12	1.32	1.55	1.67
Ngo	406	1298	1270	413	285	2.1	2.2	3.2	0.85	1.06	1.24	1.42	1.56
Nme	381	1212	1181	380	271	2.2	2.2	3.2	0.86	1.08	1.27	1.45	1.59
Osa	292	763	751	238	178	2.1	2.3	3.5	0.93	1.19	1.39	1.57	1.71
Pae	734	2453	2398	799	490	2.1	2.2	3.3	0.87	1.1	1.29	1.52	1.65
Pfu	316	901	867	283	191	2	2.3	3.4	0.93	1.14	1.33	1.5	1.65
Pgi	424	1192	1156	374	254	2.2	2.2	3.3	0.85	1.06	1.24	1.41	1.54
Pho	323	914	882	288	196	2	2.2	3.4	0.92	1.12	1.31	1.49	1.63
Spn	416	1331	1298	412	288	2.1	2.2	3.2	0.86	1.08	1.25	1.44	1.57
Rca	670	2174	2122	711	427	2.1	2.2	3.4	0.92	1.12	1.27	1.5	1.63
Rpr	214	510	504	155	100	2.3	2.3	3.4	0.91	1.11	1.27	1.44	1.57
Sce	561	1934	1889	596	402	2	2.2	3.3	0.88	1.11	1.31	1.54	1.68
Spy	403	1300	1277	404	280	2.1	2.2	3.1	0.89	1.08	1.24	1.44	1.57
Mth	430	1374	1331	428	280	2.2	2.2	3.4	0.89	1.13	1.33	1.52	1.65
Tma	338	1004	976	302	223	2.1	2.2	3.2	0.88	1.09	1.28	1.47	1.6
Tpa	207	562	555	175	124	2.2	2.3	3.1	0.86	1.03	1.21	1.42	1.55
Sty	819	3008	2951	1007	577	2.2	2.2	3.2	0.82	1.06	1.26	1.46	1.59
Ype	568	1754	1715	580	386	2.1	2.2	3.3	0.86	1.08	1.26	1.45	1.59

[a] Note: N = number of substrate, L = number of links, R = number of individual reactions or temporary substrate-enzyme complexes, E = number of enzymes, g_{in} and g_{out} = the exponents, D = diameter of the metabolic network.

2.2.2Interaction Web Database Biological Networks (IWDBNs)

The IWDB (http://www.nceas.ucsb.edu/interactionweb/resources.html) contains data sets on species interactions from several communities in different parts of the world. In a recent review, we have discussed and listed many biological networks including those contained in the IWDB.(63) Data include many types of ecological interactions: plant-pollinator, plant-frugivore, plant-herbivore, plant-ant mutualism, and predator–prey interactions. Most webs are ″bipartite networks″, which consist of two groups that are assumed to interact with species in the other group but not with species within their own group (e.g., plants and insect herbivores). Almost all data sets or webs (ecological network) are presented with an ″interaction matrix″ format (type 1 matrices), in which columns represent one group (e.g., plants) and rows represent the other group (e.g., pollinators). The exceptions to this format are predator–prey (food) webs, which are ″one-mode″ webs, represented by a symmetric matrix with all species listed in both columns and rows (type 2 matrices). In a previous work, we downloaded and transformed all matrices into .net format, which list all pairs (arcs or edges) of species (nodes) into a text file.(133) Later, we uploaded all .net files of all ecological networks to calculate numerical parameters using the MI-NODES software. This tool processes all the .net files as matrices (see the next sections). Table 2 shows a summary of all the available data sets with reference to their sources.

Table 2. Summary of Almost All Data Sets Included in the IWDB

data set[a]	habitat type	location	data type[b]	#OA[c]	#OB[c]
Anemone – Fish Networks					
1	coral reefs	Indo-Pacific	binary	10	26
Host – Parasite Networks					
2	freshwater lake	Canada	pii	7	29
3	freshwater lake	Canada	pii	10	40
4	freshwater lake	Canada	prevalence	31	144
5	river	Canada	pii	14	51
6	river	Canada	pii	17	53
7	freshwater lake	Canada	prevalence	33	97
8	freshwater reservoir	Canada	pii	6	25
45	salt marsh	USA	binary		
Plant – Ant Networks					
9	rainforest	Australia	no. visits	51	41
10	rainforest	Peru	no. visits	8	18
11	tropical forest	Costa Rica	no. visits	6	4
12	Amazon rainforest	Brazil	no. visits	16	25
Plant[a] – Herbivore Networks					
13	arid grasslands	USA	binary	54	24
				52	22
14	whole country	Finland	binary	5	64
		Britain		6	88
Plant – Pollinator Networks					
15	Andean scrub	Chile	binary	87	98
				43	62
				41	28
16	boreal forest	Canada	I. caught	12	102
17	caatinga[b]	Brazil	no. visits	13	13
18	montane forest and grassland	USA	binary	96	276
19	high-altitude desert	Canary Islands	binary	11	38
20	Alpine subarctic community	Sweden	no. visits	23	118
21	Arctic community	Canada	binary	29	86
22	heathland habitat heavily invaded by introduced plants	Mauritius Island	rates	135	74
	heathland habitat with plants removed			100	64

Table 2. Summary of Almost All Data Sets Included in the IWDB

data set[a]	habitat type	location	data type[b]	#OA[c]	#OB[c]
23	beech forest	Japan	I. caught	93	679
24	high Arctic	Canada	no. visits	32	115
25	montane forest	Australia	I. caught	42	91
26	multiple communities	Galapagos Islands	binary	106	54
27	xeric scrub	Argentina	binary	21	45
	woody riverine vegetation and xeric scrub			23	72
28	meadow	United Kingdom	F. of visits	25	79
29	Arctic community	Canada	I. caught	11	18
30	deciduous forest	USA	no. visits	13	44
31	coastal forest	Mauritius Island	no. visits	14	13
	rocky cliff and open herb community	Azores Islands		10	12
32	upland grassland	South Africa	I. caught	9	56
33	palm swamp community	Venezuela	binary	33	53
34	agricultural area	USA	binary	456	1429
35	caatinga	Brazil	binary	51	25
36	maple-oak woodland	USA	no. visits	7	32
37	peat bog	Canada	I. caught	13	34
38	evergreen montane forest	Argentina	no. visits	10	29
				9	33
				9	27
				10	29
				8	35
				8	26
				7	24
				8	27
Plant – Seed Disperser Networks					
39	forest	Papua New Guinea	no. visits	31	9
40	semideciduous tropical forest	Panama	F. removed	13	11
41	primary montane tropical rainforest	Kenya	no. visits	19	71
	secondary montante tropical rainforest			14	57
	canopy			15	71
	midstory			8	34
	understory			10	33
42	neotropical forest	Trinidad	no. visits	65	14
43	-	-	no. visits	22	20
44	temperate woodland	United Kingdom	no. visits	12	14
Predator – Prey Food Webs					
45	salt marsh	USA		binary	128
46	pine forest	New Zealand		binary	85
	pasture grassland				87
					95
					109
	tussock grassland				107
	broadleaf forest				78
					78
	pine forest	USA			78
					105
					71
					58

[a] Data set name: 1 = Ollerton et al. (2007); 2 = Aishihik Lake; 3 = Cold Lake; 4 = Lake of the Woods; 5 = McGregor River; 6 = Parsnip River; 7 = Lake Huron; 8 = Smallwood Reservoir; 9 = Blüthgen et al. (2004); 10 = Davidson et al. (1989); 11 = Davidson and Fisher (1991); 12 = Fonseca and Ganade (1996); 13 = Joern (1979); 14 = Leather (1991); 15 = Arroyo et al. (1982); 16 = Barrett and Helenurm (1987); 17 = Bezerra et al. (2009); 18 = Clements and Long (1923); 19 = Dupont et al.

(2003); 20 = Elberling and Olesen (1999); 21 = Hocking (1968); 22 = Kaiser-Bunbury et al. (2009); 23 = Kato et al. (1990); 24 = Kevan (1970); 25 = Inouye and Pyke (1988); 26 = McMullen (1993); 27 = Medan et al. (2002); 28 = Memmott (1999); 29 = Mosquin and Martin (1967); 30 = Motten (1982); 31 = Olesen et al. (2002); 32 = Ollerton et al. (2003); 33 = Ramírez and Brito (1992); 34 = Robertson (1929); 35 = Santos et al. (2010); 36 = Schemske et al. (1978); 37 = Small (1976); 38 = Vázquez and Simberloff (2002); 39 = Beehler (1983); 40 = Poulin et al. (1999); 41 = Schleuning et al. (2010); 42 = Snow and Snow (1971); 43 = Snow and Snow (1988); 44 = Sorensen (1981); 45 = Lafferty et al. (2006); 46 = Thompson and Townsend (multiple sources).

[b] Data type: pii = prevalence and intensity of infection; I. Caught = individuals caught; F. of visits = frequency of visits; F. removed = fruits removed.

[c] Number of organisms (species) with first function (#OA = number of anemone, plant, or predator species) or second function (#OB = number of fish, parasite, herbivore, pollinator, prey, or seed disperser species).

*2.2.3Spanish Financial Law Network (**SFLN**)*

The studied network is built establishing connections between two laws or legal regulations (nodes) if the time-lag is less than 1 for the same type of laws. Consequently, law-law links represent the corecurrence of different regulations in the Spanish Financial System over time, which depend in turn on social and economical conditions. The Spanish financial law recurrence network associated with the matrix **L** with elements L_{ij} was reported in previous works.(134)

2.3MI-NODES Software for the Calculation of Markov-Wiener Node Descriptors

MI-NODES (MARCH-INSIDE NOde DEScriptors) is a GUI Python/wxPython application used for the calculation of node descriptors/topological indices of nodes, subnetworks, or full networks. Actually, it should be considered as the generalization of the MARCH-INSIDE software adapted to manage any kind of complex networks (this program was originally designed to study drugs, proteins, and nucleic acid structures). MI-NODES calculates new types of node descriptors ${}^{k}C_{c}(j)$ based on Markov normalized node probabilities without a prior removal of each node to perform calculations. It also calculates Markov generalizations of different topological indices ${}^{k}TI_{c}(G)$ of class c and power k for the graph G. The tool is both Pajek and CentiBin compatible, since it is able to read networks in the following formats: .net, .dat, and .mat.

2.4MIANN Models

Let S_j be 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 ANNs to search for a nonlinear and/or linear equation with coefficients a_k, ${}^{g}b_k$, and c_0. In the particular case of a linear equation, obtained by means of a Linear Neural Network (LNN),(135) the general formula can be written as

$$S_j = \sum_{k=0}^{5} a_k \cdot W_k(j) + \sum_{g=0}^{g=Ng} \sum_{k=0}^{5} b_{gk} \cdot \left[W_k(j) - W_k(j)_{g.avg}\right] + c_0$$

$$= \sum_{k=0}^{5} a_k \cdot W_k(j) + \sum_{g=0}^{g=Ng} \sum_{k=0}^{5} b_{gk} \cdot \Delta W_k(j)_g + c_0 \qquad (5)$$

In this equation we can see the coefficients (a_k) of the Wiener-Markov node descriptors 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 subset 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.(136) However, in the present work *g* may be not only a time frame 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 the Results section.

The linear equation of the MIANN model obtained by means of LNN for MRNs is

$$S_j = \sum_{k=0}^{5} a_k \cdot W_k(j) + \sum_{k=0}^{5} b_{gk} \cdot \left[W_k(j)_{Org.avg}\right] + c_0$$

$$= \sum_{K=0}^{5} a_k \cdot W_k(j) + \sum_{g=0}^{g=Ng} \sum_{k=0}^{5} b_{gk} \cdot \Delta W_k(j)_{Org} + c_0 \quad (6)$$

The LNN model for the particular case of IWDBNs has the following formula:

$$S_j = \sum_{K=0}^{5} a_k \cdot W_k(j) + \sum_{k=0}^{5} b_{gk} \cdot \left[[W_k(j)] - W_k(j)_{Web.avg}\right] + c_0$$

$$= \sum_{k=0}^{5} a_k \cdot W_k(j) + \sum_{k=0}^{5} b_{gk} \cdot \Delta W_k(j)_{Web} + c_0 \quad (7)$$

Finally, the LNN model for the particular case of SFLN has the following formula:

$$S_j = \sum_{K=o}^{5} a_k \cdot W_k(j) + \sum_{K=0}^{5} {}^1b_k \cdot \left[W_k(j) - W_k(j)_{Year.avg}\right]$$

$$+ \sum_{k=0}^{5} {}^2b_k \cdot \left[W_k(j) - W_k(j)_{Law.avg}\right] + c_0$$

$$= \sum_{k=0}^{5} a_k \cdot W_k(j) + \sum_{K=0}^{5} {}^1b_{gk} \cdot \Delta W_k(j)_{Year}$$

$$+ \sum_{k=0}^{5} {}^2b_{gk} \cdot \Delta W_k(j)_{Law} + c_0 \quad (8)$$

where $W_k(j)$, $W_k(j)_{Year.avg}$, and $W_k(j)_{Law.avg}$ are the Wiener-Markov node descriptor parameters of a given *j*-th Law and the average of these parameters for the given year (Year.avg) or for the same type of Financial Law (Law.avg). These parameters quantify information about the Legal regulations (Laws) of a given type introduced in the Spanish legal system at a given year with respect to the previous or successive k^{th} laws approved.

In all cases, we used different statistical parameters to evaluate the statistical significance and validate the goodness-of-fit of ANN models: *n* = number of cases, Specificity, and Sensitivity of both training and external validation series.(137)

3 Results and Discussion

3.1MIANN-Wiener Models of MRNs

The study of metabolic networks is very important in biology because many applications are directly built on the use of cellular metabolism.(138, 139) Biotechnologists modify the cells and use them as cellular factories to produce antibiotics, industrial enzymes, antibodies, etc. In biomedicine, it is possible to cure metabolic diseases through a better understanding of the metabolic mechanisms and to control infections by making use of the metabolic differences between human beings and pathogens.(140) For example, the network topology-based approach has been used to uncover shared mechanisms in the study of disease comorbidity.(141) We carried out a Principal Component Analysis (PCA) of this data set (see Figure 2). We were able to explain 80% of all variance with only two principal components (pc). The first pc1, with an eigenvalue = 6.25, explains 48.1% of the variance, and the second component pc2, with an eigenvalue = 4.11, explains 31.6% of the variance. A third component pc3 was able to explain only 8.7% of the variance; consequently we discarded it. The results of this PCA are important to show that the new $W_k(j)$ indices codify useful structural information that is not trivially correlated with the information codified by other parameters. The PCA demonstrates that, in the case MRNs, the new Markov-Wiener indices codify different information compared with the classic ones.

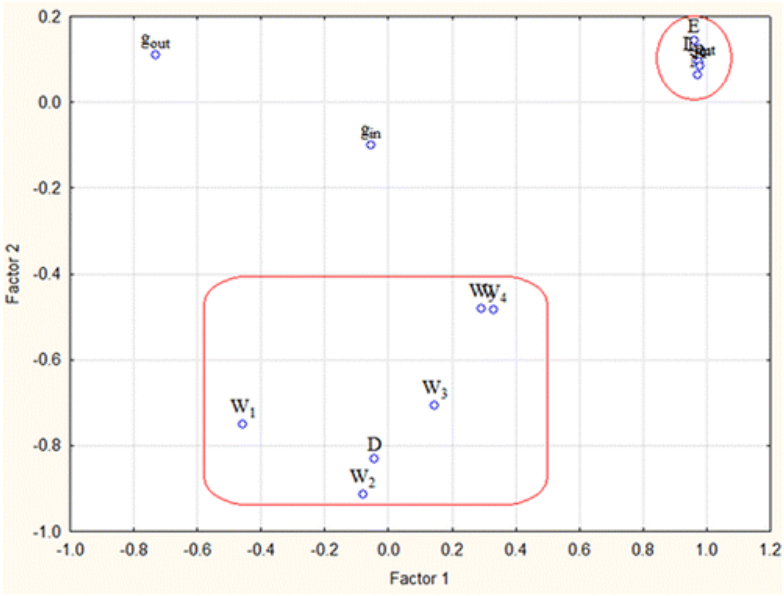
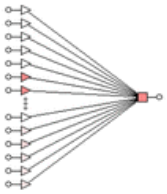
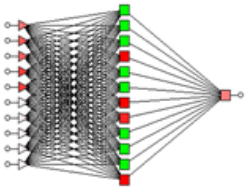


Figure 2. PCA of $W_k(j)$ values vs some classic parameters of MRNs of 43 organisms.

Jeong et al.(132) showed that, despite significant variation in their individual constituents and pathways, metabolic networks have the same topological scaling properties and show striking similarities to the inherent organization of complex nonbiological systems. In any case, many pathways are not totally confirmed experimentally but have been computationally deduced using protein or gene alignment techniques. The idea follows more or less the following scheme: similar proteome → similar enzymes → similar metabolome. On the other hand, the experimental determination of the full metabolome including each metabolite and metabolite biotransformation pathways is not always an easy task. All these aspects determine the necessity of alignment-free techniques to assess network connectivity quality in existing models of metabolic pathway networks. Here we developed different MIANN models based on W_k values to predict correct connectivity patterns of nodes in MRNs of 43 organisms belonging to different domains of the tree of life.

As seen in Table 3, 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 (see details in the Supporting Information SM1). The models were obtained using as input 15 descriptors: 5 Markov-Wiener node descriptors $W_k(j)$, 5 averages $W_k(j)_{g.avg}$, and 5 deviations – $\Delta W_k(j)_g$. The results obtained using the computer program STATISTICA show that the Multilayer Perceptron (MLP)(142) method fails to generate good prediction models, since it presents values of Specificity and Sensivity close to 50%. On the other hand, the LNN based on 15 descriptors (LNN 15:15–1:1) is able to classify correctly 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.

Table 3. MIANN Models of Metabolic Reaction Networks (MRNs)[a]

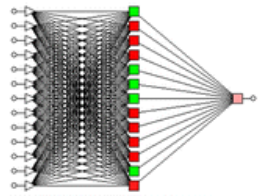
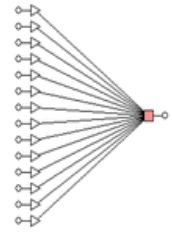
ANN	Li	Li = 1	Li = 0	%	Pr.	%	Li = 1	Li = 0
LNN 15:15-1:1 	Li = 1	7276	1985	78.1	Sn	77.9	21917	6156
	Li = 0	2044	7066	78.1	Sp	77.6	6227	21329
MLP 2:2-11-1:1 	Li = 1	4669	4559	50.1	Sn	49.7	13990	13856
	Li = 0	4651	4492	49.6	Sp	49.6	14154	13629

[a] Pr. = Parameter, Sp = Specificity, Sn = Sensitivity. Columns: Observed classifications; Rows: Predicted classifications; MLP = Multilayer Perceptron; LNN = Linear Neural Network.

3.2MIANN-Wiener Models of IWDBNs

We tested different MIANN models with linear (LNN) and nonlinear (ANN) forms. The results are presented in Table 4, and the details can be found in the Supporting Information SM2. In the case of the IWDBNs, the best classification model is obtained 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 IWDBN's nodes with accuracy (<67%). Thus, it can be observed that, compared with the MRNs, the IWDBNs contain more complex information for the classification of the connectivity between nodes. The IWDBNs need complex classifiers such as MLPs in comparison with the MRNs that can be processed using the simpler LNNs.

Table 4. MIANN Models of the IWDB Complex Networks[a]

ANN	Li	Li = 1	Li = 0	%	Pr.	%	Li = 1	Li = 0
MLP 13:13-13-1:1 	Li = 1	4570	547	91.1	Sn	90.5	1363	194
	Li = 0	449	4346	88.8	Sp	88.1	143	1437
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

[a] Pr. = Parameter, Sp = Specificity, Sn = Sensitivity. Columns: Observed classifications; Rows: Predicted classifications; MLP = Multilayer Perceptron; LNN = Linear Neural Network

3.3MIANN-Wiener Models of SFLN

The use of network analysis methods in social sciences began in 1930 and today are widely used.(143) However, the application of these methods in legal studies is still at the beginning.(144-146) The network tools can illustrate the interrelation between different laws and help to understand their consequences on the society. We have used the list of the financial laws to construct the studied network. The best models found are presented in Table 5. We tested different MIANN models with linear (LNN) and nonlinear (ANN) forms. These MIANN models behave like time series embedded within a complex network. This is due to the fact it predicts the recurrence of the Spanish law system to a financial regulation of class c when the social and economical conditions change at time t_{i+1} when a known class of regulation has been used in the past at time t_i. The best model correctly reconstructed the network of the historical record for the Spanish financial system with high Accuracy, Specificity, and Sensitivity (see Table 5). Detailed results for each case (including codes, classification, probability, and node descriptors values) are given in the Supporting Information SM3. In the case of the SFLN, there is no clear difference of Accuracy, Specificity, and Sensitivity 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. However, to be more certain of this choice, we decided to carry out a ROC curve analysis. The AUROC values (Area Under Receiver Operating Characteristic) and the ROC curves for three different MIANN models (MLP, LNN, and RBF-or Radial Basis Function-) are presented in Figure 3. We show separately the values for training and validation series. The values obtained confirm that the LNN model based on 14 descriptors is the best found in this case, with a correct classification of 86.2%, sensitivity of 87.4%, and specificity of 87.9%. The best RBF classifier, which is based on only one descriptor, is not able to classify the SFLN's nodes.

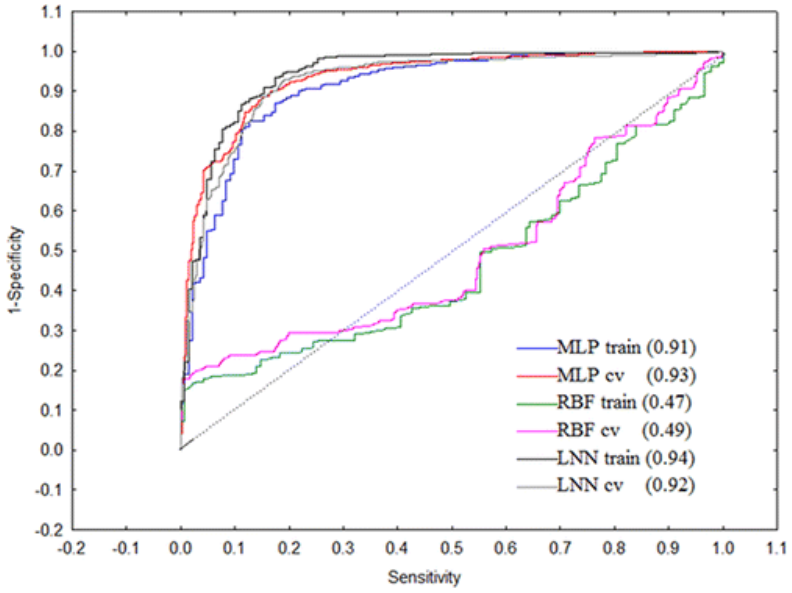
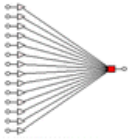
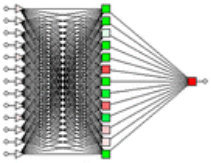


Figure 3. ROC curve analysis of the SFLN.

Table 5. MIANN Time Series Model of Spanish Financial Law Network (SFLN)[a]

ANN	Li	Li = 1	Li = 0	%	Pr.	%	Li = 1	Li = 0
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

[a] Pr. = Parameter, Sp = Specificity, Sn = Sensitivity. Columns: Observed classifications; Rows: Predicted classifications; MLP = Multilayer Perceptron; LNN = Linear Neural Network.

4 Conclusions

This work introduces a new type of node descriptors, the Markov-Wiener node descriptors of order k^{th} (W_k), higher-order analogues of the classic Wiener index, a graph invariant widely used in chemoinformatics. The new node descriptors are used to search for classification models able to discriminate the correct node connectivity patterns from the incorrect random patterns. The classifiers are obtained by using Artificial Intelligence algorithms called Artificial Neural Networks (ANNs). This mixture of Markov node descriptors and ANNs is presented as the MIANN method. The classifiers, based on LNN and MLP, showed good values of Sensitivity/Specificity (%) for the studied networks: MRNs (78/78), IWDBNs (90/88), and SFLN (86/84).

The use of the new Markov-Wiener node descriptors demonstrates that it is possible to carry out a theoretical re-evaluation of the connectivity in known complex networks (collation) as a fast alternative to the high-cost experimental re-evaluation of all the links of the studied network.

Acknowledgment

C. R. Munteanu acknowledges the Isidro Parga Pondal Program, funded by Xunta de Galicia, Spain and the European Social Fund (ESF), for partial financial support.

References

(1) Todeschini, R. and Consonni, V. Handbook of Molecular Descriptors; Wiley-VCH: 2002.
(2) Wiener, H. Structural determination of paraffin boiling points J. Am. Chem. Soc. 1947, 69, 17– 20
(3) 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– 2638
(4) 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– 1089
(5) Wiener, H. Vapor pressure–temperature relationships among the branched paraffin hydrocarbons J. Phys. Colloid Chem. 1948, 52, 425– 430
(6) 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– 2339
(7) Hosoya, H. Mathematical meaning and importance of the topological index Z Croat. Chem. Acta 2007, 80, 239– 249
(8) Consonni, V.; Todeschini, R. Molecular descriptors. In Recent advances in QSAR studies: Methods and applications; Puzyn, T.; Leszczynski, J.; Cronin, M. T. D., Eds.; Springer: 2010; Chapter 3, pp 29– 102.
(9) Harary, F. Status and contrastatus Sociometry 1959, 22, 23– 43
(10) Diudea, M. V.; Gutman, I. Wiener-type topological indices Croat. Chem. Acta 1998, 71, 21– 51
(11) Todeschini, R.; Consonni, V. Molecular Descriptors for Chemoinformatics; Wiley-VCH: Weinheim, 2009; p 1257.
(12) Balaban, A. T. From chemical graphs to 3D molecular modeling. In From chemical topology to three-dimensional geometry; Balaban, A. T., Ed.; Plenum Publishers: New York, 1997, p 420.
(13) Mandloi, M.; Sikarwar, A.; Sapre, N. S.; Karmarkar, S.; Khadikar, P. V. A comparative QSAR study using Wiener, Szeged, and molecular connectivity indices J. Chem. Inf. Comput. Sci. 2000, 40, 57– 62
(14) Lukovits, I. Decomposition of the Wiener topological index. Application to drug-receptor interactions J. Chem. Soc., Perkin Trans. 2 1988, 1667– 1671
(15) Lukovits, I. Correlation between components of the Wiener index and partition coefficients of hydrocarbons Int. J. Quantum Chem. 1992, 44, 217– 223
(16) Mendiratta, S.; Madan, A. K. Structure-activity study on antiviral 5-vinylpyrimidine nucleoside analogs using Wiener's topological index J. Chem. Inf. Comput. Sci. 1994, 34, 867– 871
(17) Agrawal, V. K.; Srivastava, R.; Khadikar, P. V. QSAR studies on some antimalarial sulfonamides Bioorg. Med. Chem. 2001, 9, 3287– 3293
(18) Sardana, S.; Madan, A. K. Application of graph theory: Relationship of molecular connectivity index, Wiener's index and eccentric connectivity index with diuretic activity MATCH 2001, 43, 85– 98
(19) Sardana, S.; Madan, A. K. Predicting anticonvulsant activity of benzamides/benzylamines: computational approach using topological descriptors J. Comput.-Aided Mol. Des. 2002, 16, 545– 550
(20) Gupta, S.; Singh, M.; Madan, A. K. Application of graph theory: Relationship of eccentric connectivity Index and Wiener's Index with anti-inflammatory activity J. Math. Anal. Appl. 2002, 266, 259– 268
(21) Bajaj, S.; Sambi, S. S.; Madan, A. K. Predicting anti-HIV activity of phenethylthiazolethiourea (PETT) analogs: computational approach using Wiener's topochemical index J. Mol. Struct. 2004, 684, 197– 203
(22) Bajaj, S.; Sambi, S. S.; Madan, A. K. Topological models for prediction of anti-HIV activity of acylthiocarbamates Bioorg. Med. Chem. 2005, 13, 3263– 3268
(23) Bajaj, S.; Sambi, S. S.; Madan, A. K. Topological models for prediction of anti-inflammatory activity of N-arylanthranilic acids Bioorg. Med. Chem. Lett. 2004, 12, 3695– 3701
(24) Bajaj, S.; Sambi, S. S.; Madan, A. K. Topochemical models for prediction of anti-tumor activity of 3-aminopyrazoles Chem. Pharm. Bull. (Tokyo) 2005, 53, 611– 615

(25) Kumar, V.; Madan, A. K. Application of graph theory: prediction of glycogen synthase kinase-3 beta inhibitory activity of thiadiazolidinones as potential drugs for the treatment of Alzheimer's disease Eur. J. Pharm. Sci. 2005, 24, 213– 218
(26) Lather, V.; Madan, A. K. Topological model for the prediction of MRP1 inhibitory activity of pyrrolopyrimidines and templates derived from pyrrolopyrimidine Bioorg. Med. Chem. Lett. 2005, 15, 4967– 4972
(27) Bornholdt, S.; Schuster, H. G. Handbook of Graphs and Complex Networks: From the Genome to the Internet; WILEY-VCH: Wheinheim, 2003.
(28) Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D. U. Complex networks: Structure and dynamics Phys. Rep. 2006, 424, 175– 308
(29) Dehmer, M.; Emmert-Streib, F. Analysis of complex networks: from biology to linguistics; Wiley-Blackwell: Weinheim, 2009; p 462.
(30) Newman, M. The structure and function of complex networks SIAM Rev. 2003, 45, 167– 256
(31) Thomas, S.; Bonchev, D. A survey of current software for network analysis in molecular biology Hum. Genomics 2010, 4, 353– 360
(32) Bonchev, D.; Buck, G. A. From molecular to biological structure and back J. Chem Inf. Model. 2007, 47, 909– 917
(33) Bonchev, D.; Rouvray, D. H. Complexity in Chemistry, Biology, and Ecology; Springer Science+Business Media, Inc.: New York, 2005.
(34) Bonchev, D. Complexity analysis of yeast proteome network Chem. Biodivers. 2004, 1, 312– 326
(35) Bonchev, D. On the complexity of directed biological networks SAR QSAR Environ. Res. 2003, 14, 199– 214
(36) Gonzalez-Diaz, H. QSAR and complex networks in pharmaceutical design, microbiology, parasitology, toxicology, cancer, and neurosciences Curr. Pharm. Des. 2010, 16, 2598– 2600
(37) Gonzalez-Diaz, H. Network topological indices, drug metabolism, and distribution Curr. Drug. Metab 2010, 11, 283– 284
(38) Vina, D.; Uriarte, E.; Orallo, F.; Gonzalez-Diaz, H. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors Mol. Pharmaceutics 2009, 6, 825– 835
(39) Duardo-Sanchez, A.; Patlewicz, G.; González-Díaz, H. A review of network topological indices from chem-bioinformatics to legal sciences and back Curr. Bioinf. 2011, 6, 53– 70
(40) Puzyn, T.; Leszczynski, J.; Cronin, M. T. D. Recent Advances in QSAR Studies: Methods and applications; Springer: London, 2010; p 423.
(41) González-Díaz, H.; Munteanu, C. R. Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks; Transworld Research Network: Kerala, India, 2010.
(42) González-Díaz, H.; González-Díaz, Y.; Santana, L.; Ubeira, F. M.; Uriarte, E. Proteomics, networks and connectivity indices Proteomics 2008, 8, 750– 778
(43) Gonzalez-Diaz, H.; Duardo-Sanchez, A.; Ubeira, F. M.; Prado-Prado, F.; Perez-Montoto, L. G.; Concu, R.; Podda, G.; Shen, B. 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
(44) Gonzalez-Diaz, H.; Prado-Prado, F.; Garcia-Mera, X.; Alonso, N.; Abeijon, P.; Caamano, O.; Yanez, M.; Munteanu, C. R.; Pazos, A.; Dea-Ayuela, M. A.; Gomez-Munoz, M. T.; Garijo, M. M.; Sansano, J.; Ubeira, F. M. 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– 1718
(45) Rodriguez-Soca, Y.; Munteanu, C. R.; Dorado, J.; Pazos, A.; Prado-Prado, F. J.; Gonzalez-Diaz, H. Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions J. Proteome Res. 2010, 9, 1182– 1190
(46) Gonzalez-Diaz, H.; Romaris, F.; Duardo-Sanchez, A.; Perez-Montoto, L. G.; Prado-Prado, F.; Patlewicz, G.; Ubeira, F. M. 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– 2764
(47) Gonzalez-Diaz, H.; Prado-Prado, F. J.; Garcia-Mera, X.; Alonso, N.; Abeijon, P.; Caamano, O.; Yanez, M.; Munteanu, C. R.; Pazos Sierra, A.; Dea-Ayuela, M. A.; Gomez-Munoz, M. T.; Garijo, M. M.; Sansano, J.; Ubeira, F. M. 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. 2010, 10, 1698– 1718

(48) Munteanu, C. R.; Vazquez, J. M.; Dorado, J.; Sierra, A. P.; Sanchez-Gonzalez, A.; Prado-Prado, F. J.; Gonzalez-Diaz, H. Complex network spectral moments for ATCUN motif DNA cleavage: first predictive study on proteins of human pathogen parasites J. Proteome Res. 2009, 8, 5219– 5228
(49) Concu, R.; Dea-Ayuela, M. A.; Perez-Montoto, L. G.; Bolas-Fernandez, F.; Prado-Prado, F. J.; Podda, G.; Uriarte, E.; Ubeira, F. M.; Gonzalez-Diaz, H. Prediction of enzyme classes from 3D structure: a general model and examples of experimental-theoretic scoring of peptide mass fingerprints of Leishmania proteins J. Proteome Res. 2009, 8, 4372– 4382
(50) Aguero-Chapin, G.; Varona-Santos, J.; de la Riva, G. A.; Antunes, A.; Gonzalez-Villa, T.; Uriarte, E.; Gonzalez-Diaz, H. 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– 2128
(51) Santana, L.; Gonzalez-Diaz, H.; Quezada, E.; Uriarte, E.; Yanez, M.; Vina, D.; Orallo, F. Quantitative structure-activity relationship and complex network approach to monoamine oxidase a and B inhibitors J. Med. Chem. 2008, 51, 6740– 6751
(52) Gonzalez-Diaz, H.; Prado-Prado, F.; Ubeira, F. M. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach Curr. Top. Med. Chem. 2008, 8, 1676– 1690
(53) Aguero-Chapin, G.; Gonzalez-Diaz, H.; de la Riva, G.; Rodriguez, E.; Sanchez-Rodriguez, A.; Podda, G.; Vazquez-Padron, R. I. 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– 448
(54) Aguero-Chapin, G.; Antunes, A.; Ubeira, F. M.; Chou, K. C.; Gonzalez-Diaz, H. Comparative study of topological indices of macro/supramolecular RNA complex networks J. Chem. Inf. Model. 2008, 48, 2265– 2277
(55) González-Díaz, H.; Vilar, S.; Santana, L.; Uriarte, E. Medicinal chemistry and bioinformatics – current trends in drugs discovery with networks topological indices Curr. Top. Med. Chem. 2007, 7, 1025– 1039
(56) Ramos de Armas, R.; González-Díaz, H.; Molina, R.; Uriarte, E. Markovian backbone negentropies: molecular descriptors for protein research. I. Predicting protein stability in Arc repressor mutants Proteins 2004, 56, 715– 723
(57) González-Díaz, H.; Marrero, Y.; Hernandez, I.; Bastida, I.; Tenorio, E.; Nasco, O.; Uriarte, E.; Castanedo, N.; Cabrera, M. A.; Aguila, E.; Marrero, O.; Morales, A.; Perez, M. 3D-MEDNEs: an alternative ″in silico″ technique for chemical research in toxicology. 1. prediction of chemically induced agranulocytosis Chem. Res. Toxicol. 2003, 16, 1318– 1327
(58) Berca, M. N.; Duardo-Sanchez, A.; González-Díaz, H.; Pazos, A.; Munteanu, C. R. Markov entropy for biology, parasitology, linguistic, technology, social and law networks. In Complex Network Entropy: From Molecules to Biology, Parasitology, Technology, Social, Legal, and Neurosciences; González-Díaz, H.; Prado-Prado, F. J.; García-Mera, X., Eds.; Transworld Research Network: Kerala, India, 2011; Chapter 10, pp 127– 142.
(59) Aguiar-Pulido, V.; Seoane-Fernańdez, J. A.; Freire-Veiga, A. M.; Gonzalez-Diaz, H.; Duardo-Sanchez, A.; Dorado, J.; Pazos, A.; Munteanu, C. R. New Markov-Randic Centralities for Computational methods of Biology, Parasitology, Technology, Social and Law networks. Proceedings of ICCMSE 2010, Kos, Greece. AIP Conference Proceedings, Melville, NY, USA, accepted (2014).
(60) González-Díaz, H.; Riera-Fernández, P.; Pazos, A.; Munteanu, C. R. The Rücker-Markov invariants of complex bio-systems: applications in parasitology and neuroinformatics Biosystems 2013, 111, 199– 207
(61) Riera-Fernandez, P.; Munteanu, C. R.; Dorado, J.; Martin-Romalde, R.; Duardo-Sanchez, A.; Gonzalez-Diaz, H. 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– 337
(62) Gonzalez-Diaz, 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. Inf. Model. 2012, 52, 3331– 3340
(63) Riera-Fernández, P.; Munteanu, C. R.; Pedreira-Souto, N.; Martín-Romalde, R.; Duardo-Sanchez, A.; González-Díaz, H. Definition of Markov-Harary invariants and review of classic topological indices and databases in biology, parasitology, technology, and social-legal networks Curr. Bioinf. 2011, 6, 94– 121
(64) Estrada, E.; Delgado, E. J.; Alderete, J. B.; Jaña, G. A. Quantum-connectivity descriptors in modeling solubility of environmentally important organic compounds J. Comput. Chem. 2004, 25, 1787– 1796
(65) Besalu, E.; Girones, X.; Amat, L.; Carbo-Dorca, R. Molecular quantum similarity and the fundamentals of QSAR Acc. Chem. Res. 2002, 35, 289– 295
(66) Rincon, D. A.; Cordeiro, M. N.; Mosquera, R. A. On the electronic structure of cocaine and its metabolites J. Phys. Chem. A 2009, 113, 13937– 13942

(67) Mandado, M.; Gonzalez-Moa, M. J.; Mosquera, R. A. Chemical graph theory and n-center electron delocalization indices: a study on polycyclic aromatic hydrocarbons J. Comput. Chem. 2007, 28, 1625– 1633
(68) Gonzalez-Diaz, H.; Gonzalez-Diaz, Y.; Santana, L.; Ubeira, F. M.; Uriarte, E. Proteomics, networks and connectivity indices Proteomics 2008, 8, 750– 778
(69) Helguera, A. M.; Combes, R. D.; Gonzalez, M. P.; Cordeiro, M. N. Applications of 2D descriptors in drug design: a DRAGON tale Curr. Top. Med. Chem. 2008, 8, 1628– 1655
(70) Casanola-Martin, G. M.; Marrero-Ponce, Y.; Khan, M. T.; Ather, A.; Khan, K. M.; Torrens, F.; Rotondo, R. Dragon method for finding novel tyrosinase inhibitors: Biosilico identification and experimental in vitro assays Eur. J. Med. Chem. 2007, 42, 1370– 1381
(71) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. Virtual computational chemistry laboratory--design and description J. Comput.-Aided Mol. Des. 2005, 19, 453– 463
(72) Marzaro, G.; Chilin, A.; Guiotto, A.; Uriarte, E.; Brun, P.; Castagliuolo, I.; Tonus, F.; Gonzalez-Diaz, H. Using the TOPS-MODE approach to fit multi-target QSAR models for tyrosine kinases inhibitors Eur. J. Med. Chem. 2011, 46, 2185– 2192
(73) Vilar, S.; Estrada, E.; Uriarte, E.; Santana, L.; Gutierrez, Y. In silico studies toward the discovery of new anti-HIV nucleoside compounds through the use of TOPS-MODE and 2D/3D connectivity indices. 2. Purine derivatives J. Chem. Inf. Model. 2005, 45, 502– 514
(74) Estrada, E.; Quincoces, J. A.; Patlewicz, G. Creating molecular diversity from antioxidants in Brazilian propolis. Combination of TOPS-MODE QSAR and virtual structure generation Mol. Divers. 2004, 8, 21– 33
(75) Estrada, E.; Gonzalez, H. What are the limits of applicability for graph theoretic descriptors in QSPR/QSAR? Modeling dipole moments of aromatic compounds with TOPS-MODE descriptors J. Chem. Inf. Comput. Sci. 2003, 43, 75– 84
(76) Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; del Valle, A.; Torrens, F.; Castro, E. A. TOMOCOMD-CARDD, a novel approach for computer-aided 'rational' drug design: I. Theoretical and experimental assessment of a promising method for computational screening and in silico design of new anthelmintic compounds J. Comput.-Aided Mol. Des. 2004, 18, 615– 634
(77) Marrero-Ponce, Y.; Medina-Marrero, R.; Castro, A. E.; Ramos de Armas, R.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F. Protein quadratic indices of the "macromolecular pseudograph's α-carbon atom adjacency matrix". 1. Prediction of Arc repressor alanine-mutant's stability Molecules 2004, 9, 1124– 1147
(78) Katritzky, A. R.; Oliferenko, A.; Lomaka, A.; Karelson, M. Six-membered cyclic ureas as HIV-1 protease inhibitors: a QSAR study based on CODESSA PRO approach. Quantitative structure-activity relationships Bioorg. Med. Chem. Lett. 2002, 12, 3453– 3457
(79) Katritzky, A. R.; Perumal, S.; Petrukhin, R.; Kleinpeter, E. Codessa-based theoretical QSPR model for hydantoin HPLC-RT lipophilicities J. Chem. Inf. Comput. Sci. 2001, 41, 569– 574
(80) Vilar, S.; Cozza, G.; Moro, S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery Curr. Top. Med. Chem. 2008, 8, 1555– 1572
(81) Junker, B. H.; Koschutzki, D.; Schreiber, F. Exploration of biological network centralities with CentiBiN BMC Bioinf. 2006, 7, 219
(82) Batagelj, V.; Mrvar, A. Pajek— analysis and visualization of large networks Lect. Notes Comput. Sci. 2002, 2265, 477– 478
(83) Hill, T.; Lewicki, P. STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining; StatSoft: Tulsa, 2006; Vol. 1, p 813.
(84) Frank, E.; Hall, M.; Trigg, L.; Holmes, G.; Witten, I. H. Data mining in bioinformatics using Weka Bioinformatics 2004, 20, 2479– 2481
(85) Bisson, W. H. Editorial: Computational chemogenomics in drug design and discovery Curr. Top. Med. Chem. 2012, 12, 1867– 1868
(86) Cordeiro, M. N.; Speck-Planche, A. Editorial: Computer-aided drug design, synthesis and evaluation of new anti-cancer drugs Curr. Top. Med. Chem. 2012, 12, 2703– 2704
(87) Prado-Prado, F.; Garcia-Mera, X. Editorial: QSAR models for computer-aided drug design and molecular docking for disorders of the central nervous system and other diseases Curr. Top. Med. Chem. 2012, 12, 1731– 1733
(88) Gonzalez-Diaz, H. Quantitative studies on Structure-Activity and Structure-Property Relationships (QSAR/QSPR); 2008; Vol. 8, p 1554.
(89) Gonzalez-Diaz, H. Editorial: QSAR/QSPR models as enabling technologies for drug & targets discovery in: medicinal chemistry, microbiology-parasitology, neurosciences, bioinformatics, proteomics and other biomedical sciences Curr. Top. Med. Chem. 2012, 12, 799– 801

(90) Alderson, R. G.; De Ferrari, L.; Mavridis, L.; McDonagh, J. L.; Mitchell, J. B.; Nath, N. Enzyme informatics Curr. Top. Med. Chem. 2012, 12, 1911– 1923
(91) Bisson, W. H. Drug repurposing in chemical genomics: can we learn from the past to improve the future? Curr. Top. Med. Chem. 2012, 12, 1883– 1888
(92) Castillo-Garit, J. A.; Abad, C.; Rodriguez-Borges, J. E.; Marrero-Ponce, Y.; Torrens, F. A review of QSAR studies to discover new drug-like compounds actives against leishmaniasis and trypanosomiasis Curr. Top. Med. Chem. 2012, 12, 852– 865
(93) Cedeno, W.; Alex, S.; Jaeger, E. P.; Agrafiotis, D. K.; Lobanov, V. S. An integrated data management framework for drug discovery--from data capturing to decision support Curr. Top. Med. Chem. 2012, 12, 1237– 1242
(94) Chatterjee, A. K.; Yeung, B. K. Back to the future: lessons learned in modern target-based and whole-cell lead optimization of antimalarials Curr. Top. Med. Chem. 2012, 12, 473– 483
(95) Chen, J.; Wang, Y.; Guo, D.; Shen, B. A systems biology perspective on rational design of peptide vaccine against virus infections Curr. Top. Med. Chem. 2012, 12, 1310– 1319
(96) Cordero, F.; Beccuti, M.; Donatelli, S.; Calogero, R. A. Large disclosing the nature of computational tools for the analysis of next generation sequencing data Curr. Top. Med. Chem. 2012, 12, 1320– 1330
(97) Dave, K.; Lahiry, A. Conotoxins: review and docking studies to determine potentials of conotoxin as an anticancer drug molecule Curr. Top. Med. Chem. 2012, 12, 845– 851
(98) Dave, K.; Panchal, H. Review on chemogenomics approach: interpreting antagonist activity of secreted frizzled-related protein 1 in glaucoma disease with in-silico docking Curr. Top. Med. Chem. 2012, 12, 1834– 1842
(99) Faivre, C.; Barbolosi, D.; Iliadis, A. A new model for determining the MTD during phase-I trials in pediatric oncology Curr. Top. Med. Chem. 2012, 12, 1660– 1664
(100) Garcia, I.; Fall, Y.; Gomez, G. Review of synthesis, biological assay, and QSAR studies of HMGR inhibitors Curr. Top. Med. Chem. 2012, 12, 895– 919
(101) Jayadeepa, R. M.; Niveditha, M. S. Computational approaches to screen candidate ligands with anti- Parkinson's activity using R programming Curr. Top. Med. Chem. 2012, 12, 1807– 1814
(102) Khan, M. T.; Mischiati, C.; Ather, A.; Ohyama, T.; Dedachi, K.; Borgatti, M.; Kurita, N.; Gambari, R. Structure-based analysis of the molecular recognitions between HIV-1 TAR-RNA and transcription factor nuclear factor-kappaB (NFkB) Curr. Top. Med. Chem. 2012, 12, 814– 827
(103) Kobe, B.; Boden, M. Computational modelling of linear motif-mediated protein interactions Curr. Top. Med. Chem. 2012, 12, 1553– 1561
(104) Kramer, C.; Lewis, R. QSARs, data and error in the modern age of drug discovery Curr. Top. Med. Chem. 2012, 12, 1896– 1902
(105) Kufareva, I.; Chen, Y. C.; Ilatovskiy, A. V.; Abagyan, R. Compound activity prediction using models of binding pockets or ligand properties in 3D Curr. Top. Med. Chem. 2012, 12, 1869– 1882
(106) Lin, J. H. Target prediction of small molecules with information of key molecular interactions Curr. Top. Med. Chem. 2012, 12, 1903– 1910
(107) Luan, F.; Borges, F.; Cordeiro, M. N. Recent advances on A(3) adenosine receptor antagonists by QSAR tools Curr. Top. Med. Chem. 2012, 12, 878– 894
(108) Mortier, J.; Rakers, C.; Frederick, R.; Wolber, G. Computational tools for in silico fragment-based drug design Curr. Top. Med. Chem. 2012, 12, 1935– 1943
(109) Ortore, G.; Tuccinardi, T.; Martinelli, A. Computational studies on translocator protein (TSPO) and its ligands Curr. Top. Med. Chem. 2012, 12, 352– 359
(110)
(111) Popelier, P. New insights in atom-atom interactions for future drug design Curr. Top. Med. Chem. 2012, 12, 1924– 1934
(112) Prado-Prado, F.; Garcia-Mera, X.; Escobar, M.; Alonso, N.; Caamano, O.; Yanez, M.; Gonzalez-Diaz, H. 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– 1865
(113) Riera-Fernandez, I.; Martin-Romalde, R.; Prado-Prado, F. J.; Escobar, M.; Munteanu, C. R.; Concu, R.; Duardo-Sanchez, A.; Gonzalez-Diaz, H. 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– 960
(114) Saladino, G.; Gervasio, F. L. New insights in protein kinase conformational dynamics Curr. Top. Med. Chem. 2012, 12, 1889– 1895
(115) Sharma, N.; Ethiraj, K. R.; Yadav, M.; Nayarisseri, S. A.; Chaurasiya, M.; Vankudavath, R. N.; Rao, K. R. Identification of LOGP values and electronegativities as

structural insights to model inhibitory activity of HIV-1 capsid inhibitors - A SVM and MLR aided QSAR studies Curr. Top. Med. Chem. 2012, 12, 1763– 1774
(116) Speck-Planche, A.; Kleandrova, V. V. QSAR and molecular docking techniques for the discovery of potent monoamine oxidase B inhibitors: computer-aided generation of new rasagiline bioisosteres Curr. Top. Med. Chem. 2012, 12, 1734– 1747
(117) Van Calenbergh, S.; Pochet, S.; Munier-Lehmann, H. Drug design and identification of potent leads against mycobacterium tuberculosis thymidine monophosphate kinase Curr. Top. Med. Chem. 2012, 12, 694– 705
(118) Zhang, T.; Zhao, M.; Pang, Y.; Zhang, W.; Angela Liu, L.; Wei, D. Q. Recent progress on bioinformatics, functional genomics, and metabolomics research of cytochrome P450 and its impact on drug discovery Curr. Top. Med. Chem. 2012, 12, 1346– 1355
(119) Gonzalez-Diaz, H.; Romaris, F.; Duardo-Sanchez, A.; Perez-Montoto, L. G.; Prado-Prado, F.; Patlewicz, G.; Ubeira, F. M. 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– 2764
(120) Caballero, J.; Fernandez, M. Artificial neural networks from MATLAB in medicinal chemistry. Bayesian-regularized genetic neural networks (BRGNN): application to the prediction of the antagonistic activity against human platelet thrombin receptor (PAR-1) Curr. Top. Med. Chem. 2008, 8, 1580– 1605
(121) Duardo-Sanchez, A.; Patlewicz, G.; Lopez-Diaz, A. Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues Curr. Top. Med. Chem. 2008, 8, 1666– 1675
(122) Ivanciuc, O. Weka machine learning for predicting the phospholipidosis inducing potential Curr. Top. Med. Chem. 2008, 8, 1691– 1709
(123) Wang, J. F.; Wei, D. Q.; Chou, K. C. Drug candidates from traditional chinese medicines Curr. Top. Med. Chem. 2008, 8, 1656– 1665
(124) Gonzalez-Diaz, H.; Vilar, S.; Santana, L.; Uriarte, E. Medicinal chemistry and bioinformatics - Current trends in drugs discovery with networks topological indices Curr. Top. Med. Chem. 2007, 7, 1015– 1029
(125) Bhattacharjee, B.; Jayadeepa, R. M.; Banerjee, S.; Joshi, J.; Middha, S. K.; Mole, J. P.; Samuel, J. Review of complex network and gene ontology in pharmacology approaches: Mapping natural compounds on potential drug target colon cancer network Curr. Bioinf. 2011, 6, 44– 52
(126) Dave, K.; Banerjee, A. Bioinformatics analysis of functional relations between CNPs regions Curr. Bioinf. 2011, 6, 122– 128
(127) García, I.; Fall, Y.; Gómez, G. Trends in bioinformatics and chemoinformatics of vitamin D analogues and their protein targets Curr. Bioinf. 2011, 6, 16– 24
(128) Ivanciuc, T.; Ivanciuc, O.; Klein, D. J. Network-QSAR with reaction poset quantitative superstructure-activity relationships (QSSAR) for PCB chromatographic properties Curr. Bioinf. 2011, 6, 25– 34
(129) Prado-Prado, F.; Escobar-Cubiella, M.; García-Mera, X. Review of bioinformatics and QSAR studies of β-secretase inhibitors Curr. Bioinf. 2011, 6, 3– 15
(130) Wan, S. B.; Hu, L. L.; Niu, S.; Wang, K.; Cai, Y. D.; Lu, W. C.; Chou, K. C. Identification of multiple subcellular locations for proteins in budding yeast Curr. Bioinf. 2011, 6, 71– 80
(131) Gonzalez-Diaz, H.; Duardo-Sanchez, A.; Ubeira, F. M.; Prado-Prado, F.; Perez-Montoto, L. G.; Concu, R.; Podda, G.; Shen, B. 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
(132) Gonzalez-Diaz, H.; Arrasate, S.; Sotomayor, N.; Lete, E.; Munteanu, C. R.; Pazos, A.; Besada-Porto, L.; Ruso, J. M. MIANN models in medicinal, physical and organic chemistry Curr. Top. Med. Chem. 2013, 13, 619– 641
(133) Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; Barabasi, A. L. The large-scale organization of metabolic networks Nature 2000, 407, 651– 654
(134) Riera-Fernandez, P.; Munteanu, C. R.; Pedreira-Souto, N.; Martin-Romalde, R.; Duardo-Sanchez, A.; Gonzalez-Diaz, H. Definition of Markov-Harary invariants and review of classic topological indices and databases in biology, parasitology, technology, and social-legal networks Curr. Bioinf. 2011, 6, 94– 121
(135) Riera-Fernandez, P.; Munteanu, C. R.; Escobar, M.; Prado-Prado, F.; Martin-Romalde, R.; Pereira, D.; Villalba, K.; Duardo-Sanchez, A.; Gonzalez-Diaz, H. 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– 188
(136) Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities Proc. Natl. Acad. Sci. U. S. A. 1982, 79, 2554– 2558
(137) Box, G. E. P.; Jenkins, G. M. Time series analysis; Holden-Day: San Francisco, 1970; p 553.

(138) Hill, T.; Lewicki, P. STATISTICS Methods and Applications; StatSoft: Tulsa, 2006.
(139) Mazurie, A.; Bonchev, D.; Schwikowski, B.; Buck, G. A. Evolution of metabolic network organization BMC Syst. Biol. 2010, 4, 59
(140) Kier, L. B.; Bonchev, D.; Buck, G. A. Modeling biochemical networks: a cellular-automata approach Chem. Biodivers. 2005, 2, 233– 243
(141) Rosa da Silva, M.; Sun, J.; Ma, H. W.; He, F.; Zeng, A. P. Metabolic networks. In Analysis of biological networks; Junker, B. H.; Schreiber, F., Eds.; Wiley & Sons: NJ, 2008, pp 233– 253.
(142) Lee, D. S.; Park, J.; Kay, K. A.; Christakis, N. A.; Oltvai, Z. N.; Barabasi, A. L. The implications of human metabolic network topology for disease comorbidity Proc. Natl. Acad. Sci. U. S. A. 2008, 105, 9880– 9885
(143) Rosenblatt, F. Principles of neurodynamics; perceptrons and the theory of brain mechanisms; Spartan Books: WA, 1962.
(144) Wasserman, S.; Faust, K. Social network analysis: methods and applications; Cambridge University Press: Cambridge, 1999.
(145) Fowler, J. H.; Jeon, S. The authority of Supreme Court precedent Social Networks 2008, 30, 16– 30
(146) Duardo-Sánchez, A. Study of criminal law networks with Markov-probability centralities. In Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks; González-Díaz, H., Ed.; Bentham: Kerala, India, 2010; Chapter 12, pp 205– 212.
(147) Duardo-Sánchez, A. Criminal law networks, markov chains, Shannon entropy and artificial neural networks. In Complex Network Entropy: From Molecules to Biology, Parasitology, Technology, Social, Legal, and Neurosciences; González-Díaz, H., Ed.; Bentham: Kerala, India, 2011; Chapter 8, pp 107– 114.