Current Bioinformatics. 2015; 10 (5): 658-671

MIANN models of networks of biochemical reactions, ecosystems, and U.S. Supreme Court with Balaban-Markov indices

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

[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 11, University of the Basque Country UPV/EHU, 48940, Leioa, Spain*
[4] *IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

Abstract:
We can use Artificial Neural Networks (ANNs) and graph Topological Indices (TIs) to seek structure-property relationship. Balabans' J index is one of the classic TIs for chemo-informatics studies. We used here Markov chains to generalize the J index and apply it to bioinformatics, systems biology, and social sciences. We seek new ANN models to show the discrimination power of the new indices at node level in three proof-of-concept experiments. First, we calculated more than 1,000,000 values of the new Balaban-Markov centralities Jk(i) and other indices for all nodes in >100 complex networks. In the three experiments, we found new MIANN models with >80% of Specificity (Sp) and Sensitivity (Sn) in train and validation series for Metabolic Reactions of Networks (MRNs) for 42 organisms (bacteria, yeast, nematode and plants), 73 Biological Interaction Webs or Networks (BINs), and 43 sub-networks of U.S. Supreme court citations in different decades from 1791 to 2005. This work may open a new route for the application of TIs to unravel hidden structure-property relationships in complex bio-molecular, ecological, and social networks.

Keywords:
Artificial neural networks; Markov chains; U.S. supreme court; Complex networks; Ecosystems; Legal and social networks; Metabolomics

1. INTRODUCTION

Complex networks are present in almost all the levels of material world [1], as observed after an inspection of complex networks formed at multi scales: from networks of metabolic organic reactions inside living beings [2], passing through interactions between species in food webs and ecosystems [3], extending up to decisions in the U.S. Supreme Court [4, 5]. We make emphasis on the three previous examples because they will be subjected to closer inspection in this work. However, the number of examples is huge including also networks of Inorganic and Organic Reactions in the atmosphere [6]; Obesity, or Alcoholism spreading in social networks [7, 8]; Internet and the World Wide Web (WWW), Electric power distribution networks, *etc.* [9]. In any case, we can represent all these complex systems by a simple draw (graph) linking the parts of the system (no des) by means of edges that express the existence of ties, interactions, or some relationship between nodes.

2. REVIEW OF BALABANS' J INDEX

Network analysis may capture static or dynamic information about complex systems [10]. It is straightforward to realise, from the previous section, that we can use numeric parameters that quantify the different connectivity patterns in a graph to unravel information about hidden structure-property relationships (static or dynamic) in complex systems. Many authors call these numeric parameters as Connectivity or Topological indices (TIs) [11, 12].

TIs are numeric parameters used to describe the information about molecular structure of a molecule (molecular descriptor) or large systems (bio-molecular, social, technological, etc.) susceptible to be represented as a network or graph (G). For instance, we can define G ≡ (v, e) as a list of all pairs formed by the v^{th} vertices or nodes (e.g.; atoms, proteins, countries...) interconnected by a list of e^{th} edges, ties or relationships (chemical bonds, protein interactions, bilateral trade...). We can depict G in the plane (2D space) as a graphic representation of all these nodes (dots) interconnected by ties (edges). In any case, we can also construct graphs with vertices as points in a 3D space, and edges connecting these vertices. See for instance, the work of Komosinski and Kubiak [13] in complexity about 3D morphologies.

Consequently, TIs are numerical parameters of a graph used to quantify information about the topology (in the sense of node connectivity) and are usually graph invariants (do not depend on node labels, rotation, translations, or geometric distortion of graph edges in general). In general, TIs are called to play an important role in the characterization of the complexity of real systems that can be represented as graphs or complex networks. For instance, Diudea *et al.* [14] published a paper in the journal complexity introducing a new super index based on Shell- matrices and polynomials.

Dehmer and Mowshowitz [15] published in the same journal a generalization of entropy measures applying information measures to a graph. In Prof. Balaban's words, TIs are a convenient means of translating chemical constitution into numerical values [16]. Until now, multiple TIs have been defined and/or applied to some extend. Many of them are interrelated somehow and other TIs are more exotic. For instance, Basak and Balaban, *et al.* [17] calculated 202 molecular descriptors (TIs) for two chemical databases. The first contained 139 hydrocarbons and the second one a heterogeneous series of 1037 diverse compounds. They were able to group all these TIs into 14-18 clusters using variable cluster analysis. Correspondences between the same TIs in the two sets revealed how and why the various classes of TIs are mutually related and provide insight into what aspects of chemical structure they are expressing [17].

In any case, some of these TIs have become classic and/or TIs of reference with the passage of time. This is the case of Wiener [18-20], Hosoya [21, 22], Randic [23], Kier and Hall [24, 25], and other TIs. According to the authors [26] TIs, can belong to three generations or classes. The classes move from first-generation of TIs having the form of integer numbers with high degeneracy that limits their use to more elaborated TIs. In particular, the same professor mentioned in the first sentence introducing one TIs; which have become one of the classics and known as the Balaban's J index [16]. The J index is a function of 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. The J index also depends on Si = distance sums ca1culated as the sums over the rows or columns of the topological distance matrix **D** of the hydrogen-depleted graph G for the molecule. The formula of this classic TI is:

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

3. APPLICATIONS OF MOLECULAR INFORMATICS

Balaban's J index have been used in many chemo- informatics studies as input for many different Machine Learning (ML) algorithms; alone and/or combined with other TIs. Almost all applications of Balban's J index deal with drug discovery; in particular the prediction of drugs with higher biological activity and/or low toxicity. For instance, Sharma *et al.* [27]; synthesized new carboxylic acid ethyl esters and evaluated the *in vitro* antimicrobial and anticancer activity of these compounds. In a chemoinformatic study they demonstrated the importance of Balaban's J index and log P to describe the antimicrobial activity of these compounds.

Thakur *et al.* [28]; carried out a QSAR study on benzenesulphonamide carbonic anhydrase inhibitors using Balaban's J index. The regression analysis has shown that even in mono-parametric regression this index gave excellent results. Moreover, the combination of J with the first-order Randic connectivity index $^{1}\chi$ improved the results obtained. Krawczuk *et al.* [29]; used the J index and the electro- topological index for the prediction of retention data of polychlorinated biphenyls. J index has been used also as index of reference to test the performance of new TIs. In any case, a large number of applications of J index is in medicinal chemistry. Yadav, Kumar, and De Clercq [30] *et al.*, synthesized and evaluated the *in vitro* antimicrobial activity of new sulphonic acids. They compared the results of one-target *vs* multi-target models for these compounds. The multi-target model based on Balaban index (J), a LUMO parameter, and second order valence connectivity index ($^{2}\chi_{v}$) was very useful to describe the antimicrobial activity of synthesized compounds. Naik, Dubey, and Kumar [31], developed new predictive models for epipodophyllotoxin derivatives. Epipodophyllotoxins are important anticancer drugs used in chemotherapy for various types of cancers. The model was obtained with descriptors such as solvent- accessible surface area, heat of formation, Balaban index, number of atom classes, and sum of E-state index of atoms. Fernandes [32] predicted the activity of forty-three pyrazinoates against *M tuberculosis* ATCC 27294, using Balaban index (1) and other parameters as inputs for a gene tic algorithm function and partial least squares regression (WOLF 5.5 program).

Panaye, Doucet, and Devillers, *et al.* [33]; developed compared decision trees *vs* support vector machine (SVM) for classification of androgen receptor ligands. They predicted relative binding affinity (RBA) to a large set of about 200 chemicals with descriptors calculated from CODESSA software including hydrophobicity parameter (logP), Balaban index, and other descriptors. Zavrsnik, Muratović, and Spirtović [34]; reported the synthesis of 4- arylaminocoumarin derivatives with antimycotic effects and a QSAR study of this activity with, Balaban J index, Wiener W index, and other TIs and physicochemical properties.

Ma, Chen, and Yang [35]; used the Balaban index and other molecular descriptors to model blood-brain barrier (BBB) penetration of different compounds. One aspect of special relevance for this work is the use of Balaban's J index as input for Artificial Neural Networks (ANNs). Dashtbozorgi and Golmohammadi [36], used Multiple Linear Regression (MLR) and ANNs to model water-to-wet butyl acetate partition coefficient of 76 organic solutes using Balaban index (1) together with the Kier and Hall index of order 2 ($^{2}\chi$), and other indices. Jalali-Heravi and Fatemi [37]; developed ANN model using Balaban index and other parameters to predict the thermal conductivity detection response factors of 110 organic compounds.

In other line of thinking, some author have reported applications of Balban's J index in mathematical and/or physical-chemistry; including the generalization of this index to create other TIs (called Balaban type parameters). Dehmer *et al.* [38] compared the discriminative power for graphs of a new information index *vs* the J values for benchmark dataset of nearly 12 million exhaustively generated, non-isomorphic, and unweighted graphs. In this sense, Basak *et al. [39];* reviewed the use of mathematical structural invariants in analyzing combinatorial libraries.

Ratkiewicz and Truong [40]; reported a new method for automatic generation of mechanisms of reactions of complex systems using Balaban, Schulz, Connectivity and other TIs. Rastija and Medié-Sarié [41] studied antioxidant activity of wine polyphenols with Balaban index, Balaban-type index and other descriptors. Randié and Pompe [42]; 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 1/1.

4. MIANN MODELS

Dehmer *et al.* [43]; noted that many TIs of complex networks have been developed without giving a proper proof of their potential applications. They also talk about the high interest in the development of software packages to ca1culate TIs. We can use computer software like CENTIBIN [44], PAJEK [45], and QUACN [43, 46] which can be used to calculate TIs of complex networks.

Specifically, different authors have reported, and/or used different types of TIs that may be considered variations or generalizations of the J index to some extent. For instance, Randić & Pompe [42]; introduced several variable molecular descriptors, derived from the distance matrix and the "reversed" distance matrix. This includes 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. Balaban type indices: J, Jz, Jm, Jv, Jc, and Jp have been used to predict the supra-molecular complexing ability of a sulfonamides [47].

González-Díaz *el al.* [48]; developed the software MARCH-INSIDE. The software MI uses Markov Chains theory to ca1culate TIs of order k (TI_k) inside very different structures in many file formats such as SMILE .txt and .mol file of drugs, .pdb files of proteins, or .ct files for RNAs. In any case, MI cannot upload formats of Complex Networks (.mat, .net, .dat, *etc.).* Accordingly, a new software called MI-NODES (MARCH-INSIDE NOde DEScriptors) has been proposed. MI-NODES can upload many formats of complex networks and calculate several TI_k values for these networks. MI-NODES incorporates node-node transition Markov probabilities (p_{ij}) inside the formula of classical TIs. In so doing, we can ca1culate new versions of classic TI(G) of a full graph G based on Markov Chains $TI_k(G)$ and the respective node centralities $TI_k(i)$ [49-52].

MI indices can be as inputs in Linear Discriminant Analysis (LDA) executed in STATISCA [53] or ML methods developed in WEKA program [54]. Especially, we can use Artificial Neural Networks (ANNs); which is very powerful bio- inspired Artificial Intelligence (Al) algorithms. This approach is called MIANN (MI & ANN models). We have reviewed the MIANN in a recent work. In this paper we focused on theoretical basis, development of web servers and, the uses in molecular sciences [55]. In any case, almost al MIANN models use the original MI software and focus on chemical or bio-molecular systems. Conversely, the use of MI-NODES to seek MIANN models is relatively a new direction.

Taking into consideration all the aspects discussed in the introduction we decided to perform the following work. In this work we generalized the Balaban' s J index in a different direction, implement the ca1culation of these TIs in MI- NODES, and seek new MIANN models to show in proof-of- concept experiments to demonstrate the discrimination power of the new indices.

4.1. Theory of MIANN Models

In many complex networks we can found certain uncertainty on the assignation of links between nodes due to errors or existence of contradictory information obtained by different experimental ways. Let $\mathbf{L}_i$ be the connectivity pattern (set of all walks of length k from the i^{th} node to the remnant *n* -1 nodes in the network) we can try to seek a model to discriminate the observed (correct) $\mathbf{L}_i$ of all nodes in real networks *vs* no correct connectivity patterns. The output of this type of models may be a real value function Sj which scores the reliability of a sub-graph Li; Li = 1 when the connections are correct others (real or correct Li) in the network or Li = 0 for nodes that we know a priori that present an incorrect connectivity pattern. We can use different TIs and/or no de centralities to describe numerically different connectivity patterns **L**i (correct or incorrect patterns in a give network). Next, we can use these TIs as inputs of ANN algorithms in order to search for non-linear and/or linear models able to predict the correct **L**is. In LNN model we can write the general formula of this type of MIANN models as follows:

$$S_j = \sum_{k=0}^{5} a_k \cdot TI_k(j) + \sum_{g=0}^{g=Ng} \sum_{k=0}^{5} b_{gk} \cdot \left[TI_k(j) - TI_k(j)_{g-avg}\right] + c_0$$
$$= \sum_{k=0}^{5} a_k \cdot TI_k(j) + \sum_{g=0}^{g=Ng} \sum_{k=0}^{5} b_{gk}.\Delta TI_k(j)_g + c_0 \quad (2)$$

In these equations, TIk are node centralities based on topological indices ca1culated by software MI-NODES based on the MARCH-INSIDE algorithm. However, we can use any TI calculated with other software. The coefficients (a_k) quantify the influence of the centralities of nodes used as input. The coefficients ($^{g}b_k$) quantify the effect of the deviation of the TIs of a given no de with respect to the average value of the TIs of a sub-set of nodes that obey a given condition in a network of reference correctly constructed. The deviation operators are ΔTIk (j)g = [TIk (j) - TIk (j)g.avg]; with TIk (j)g.avg equal to the expected value or average (avg) of TIk(j) for a sub-set or group (g) of nodes similar to Box- Jenkins' ARIMA models [56]. Specifically, we can write the linear equation of the MIANN model obtained by LNN analysis for MRNs is as follows:

$$S_i = \sum_{k=0}^{5} a_k \cdot TI_k(i) + \sum_{k=0}^{5} b_{gk} \cdot \left[TI_k(i) - TI_k(i)_{Org.avg}\right] + c_0$$
$$= \sum_{k=0}^{5} a_k \cdot TI_k(i) + \sum_{g=0}^{g=Ng} \sum_{k=0}^{5} b_{gk} \cdot \Delta TI_k(i)_{Org} + c_0 \quad (3)$$

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

$$S_i = \sum_{k=0}^{5} a_k \cdot TI_k(i) + \sum_{k=0}^{5} b_{gk} \cdot \left[TI_k(i) - TI_k(i)_{Web.avg}\right] + c_0$$
$$= \sum_{k=0}^{5} a_k \cdot TI_k(i) + \sum_{k=0}^{5} b_{gk} \cdot \Delta TI_k(i)_{Web} + c_0 \quad (4)$$

Last, the LNN model for the particular case of USSCN has the following formula:

$$S_i = \sum_{k=0}^{5} a_k \cdot TI_k(i) + \sum_{k=0}^{5} {}^{1}b_k \cdot \left[TI_k(i) - TI_k(i)_{Year.avg}\right]$$
$$+ \sum_{k=0}^{5} {}^{2}b_k \cdot \left[TI_k(j) - TI_k(j)_{Citing.avg}\right] + c_0$$
$$= \sum_{k=0}^{5} a_k \cdot TI_k(i) + \sum_{k=0}^{5} {}^{1}b_k \cdot \Delta TI_k(i)_{Year}$$
$$+ \sum_{k=0}^{5} {}^{2}b_{gk} \cdot \Delta TI_k(i)_{Citing} + c_0 \quad (5)$$

where, the $J_k(j)$ values are the Balaban-Markov centrality parameters of a given j-th judicial cases resolved by the U.S. Supreme Court. Whereas, the $J_k(j)_{Year.avg}$ and $J_k(j)_{Citing.avg}$ values are the average of these parameters for all the cases in this court in the given year (Year.avg) or for the cases citing j^{th} (Citing.avg). In addition, we can use statistical parameters of both train and external validation series to assess the goodness-of-fit of ANN models: *n* = number of cases, Specificity, and Sensitivity [57].

5. NEW MIANN MODELS WITH J_K INDICES

In the present work, we introduce for the first time a new type of Balaban type indices called Balaban-Markov. The global Balaban-Markov MG) indices are useful to study the full graph G of a molecular or complex network system in general. The respective no de centralities $J_k(i)$ are expected to be useful in the study complex networks at node level. The calculation of the new indices is implemented in MI- NODES. Subsequently, we applied the MIANN strategy to study complex bio-molecular complex networks. In addition, we studied ecological and social-legal systems as well to illustrate the uses of the new parameter in higher scales. In closing, we developed three proof-of-concept experiments to i11ustrate the potentialities of the new parameters. In each experiment, we seek new MIANN models for the first time useful to discriminate correct from unreliable connectivity patterns in complex networks. In the first experiment, we found new MIANN-Balaban models for Metabolic Reactions of Networks (MRNs) for 43 organisms (bacteria, yeast, nematode and plants). In the second experiment, we developed MIANN-Balaban for 73 Biological Interaction Webs or Networks (BINs). The biological interactions or relationships present in these networks include: prey-hunter, parasite-host, plant-seed disperser, anemone-clown fish species and others. In the third experiment, we found MIANN-Balaban models for >40 networks relevant for Legal sciences. Each one of these networks (5KCNs) represent one slot of 5000 (5K) cites to 1000-5000 cases of the U.S. Supreme court in different decades (1700-2006). Despite the differences between the different networks the workflow used in all the experiments is essentially the same (see Fig. 1). It is straightforward to realize from Fig. (1) that the given steps are essentially the same used in QSAR/QSPR analysis of molecular systems in chemoinformatics. The datasets used to develop the new MIANN models are the following.

5.1. Data 1: Metabolic Reaction Networks (MRNs)

These MRNs were downloaded in a zipped ASCII file directly from Barabasi's group web (http://www.nd.edu/~net works/resources.htm). The metabolic networks data was obtained from a previous work after Jeong *et al.* (see details there) [2]. In Table 1 we report the values of the $J_k(G)$ indices for the MRN s of different organisms studied in this work. We also depict the values of classic parameters of these MRNs that have been reported by other authors. In the next sections we shall discuss the relationships between $J_k(G)$ indices and these classic parameters.

5.2. Data 2: Biological Interaction Networks (BINs)

In a recent review we discussed many biological interaction webs or networks (BINs) including those contained in the IWDB [58]. In a previous work, we downloaded all matrices compiled in the IWDB and transformed them into BINs in .net format. This format lists all pairs (arcs or edges) of species (nodes) into a text file [59]. The IWDB (http://www.nceas.ucsb.edu/interactionweb/resources.html) currently collects datasets about food webs for different species and ecosystems. In Table 2 we can find a summary of all the available datasets; see a list of references to original sources in our previous work [59]. The full list of reference is too large to be cited here.

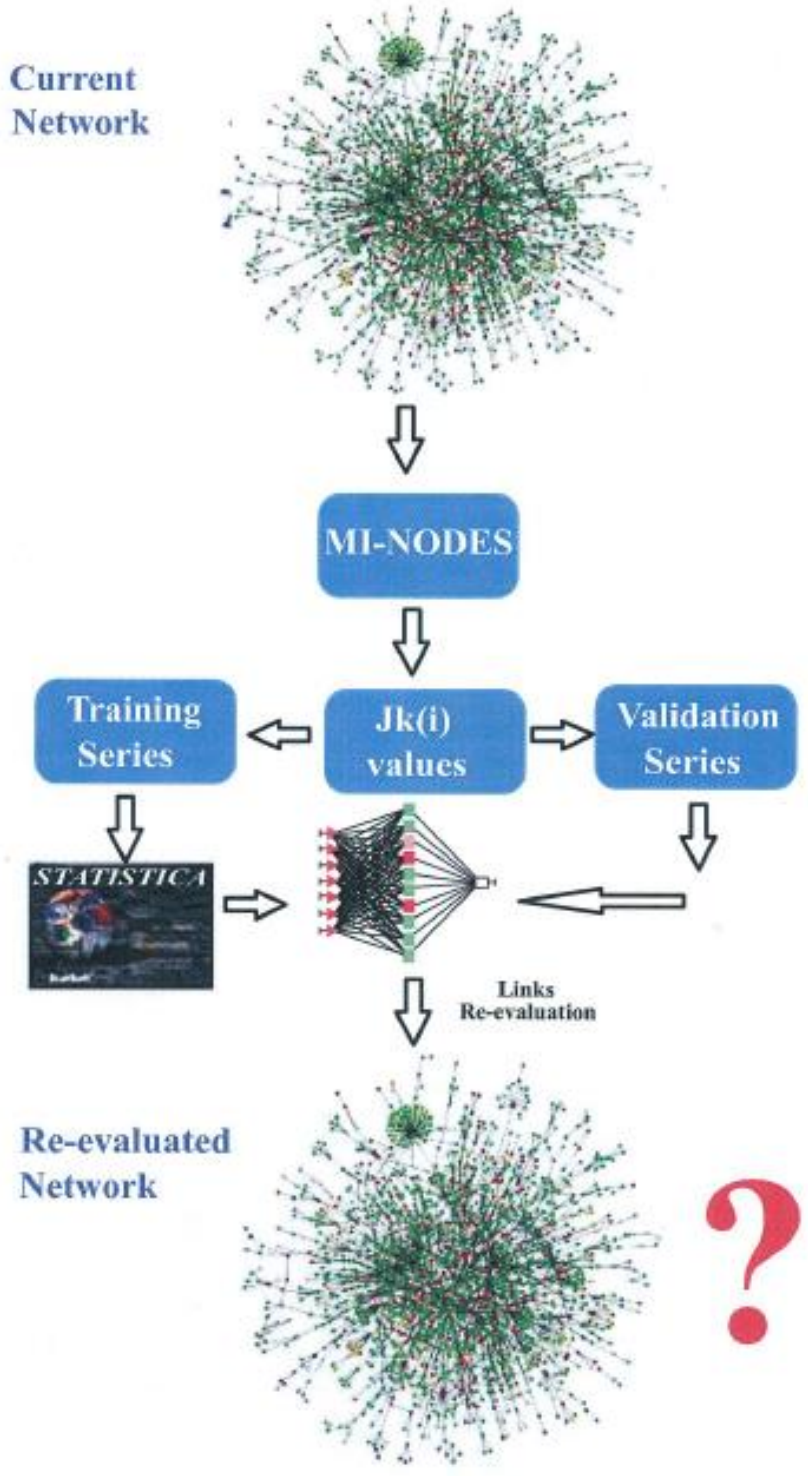


Fig. (1). General workflow used in this work.

Table 1. Average values $J_k(i)_{org.avg}$ *vs* classic parameters of MRNs of some organisms.

MRN	J_o	J_1	J_2	J_3	J_4	J_5	N	L_{in}	L_{out}	R	E	g_{in}	g_{out}	D
AA	3289.9	3233.4	1863.0	1117.6	774.8	677.8	419	1278	1249	401	285	2.10	2.20	3.30
AB	2933.9	2869.2	1625.4	975.6	682.6	634.4	395	1202	1166	380	271	2.10	2.20	3.20
AG	4234.2	4165.9	2432.1	1471.4	1028.9	910.0	496	1527	1484	486	299	2.20	2.20	3.50
AP	1514.3	1474.5	806.4	458.4	315.1	328.9	204	588	575	178	135	2.20	2.20	3.20
AT	2324.1	2284.3	1343.7	822.8	580.1	508.9	302	804	789	250	185	2.10	2.30	3.50
BB	1085.8	1057.7	581.7	336.6	235.8	225.1	187	442	438	140	106	2.30	2.40	3.00
BS	6513.1	6268.7	3314.5	1957.0	1331.8	1204.5	785	2794	2741	916	516	2.20	2.10	3.30
CA	4018.9	3938.0	2220.9	1313.5	904.5	794.6	494	1624	1578	511	344	2.10	2.20	3.30
CE	3707.9	3642.9	2067.0	1224.1	842.8	757.7	462	1446	1418	450	295	2.10	2.20	3.30
Cl	2834.7	2774.5	1595.4	959.9	668.4	584.8	380	1142	1115	359	254	2.10	2.30	3.20
CL	2821.7	2742.7	1479.3	839.2	561.1	556.9	389	1097	1062	333	231	2.10	2.20	3.30
CQ	1238.6	1193.7	630.6	352.4	242.9	253.5	194	401	391	134	84	2.20	2.30	3.40
CT	1491.2	1466.7	904.3	576.3	419.2	374.0	215	479	462	158	94	2.20	2.40	3.50
CY	4234.7	4089.9	2101.2	1181.1	788.0	804.8	546	1782	1746	570	370	2.00	2.20	3.30
DR	6805.0	6614.5	3609.3	2147.1	1480.5	1345.6	815	2870	2811	965	557	2.20	2.10	3.30
EC	6596.4	6445.0	3629.6	2185.4	1509.8	1363.7	778	2904	2859	968	570	2.20	2.10	3.20
EF	2987.0	2936.2	1699.0	1010.1	693.0	608.0	386	1244	1218	382	281	2.10	2.20	3.10
EN	2870.5	2814.5	1585.1	937.1	645.8	576.2	383	1095	1081	339	254	2.10	2.20	3.30
HI	4161.9	4019.0	2098.8	1239.7	863.2	855.3	526	1773	1746	597	361	2.10	2.30	3.20
HP	2833.8	2759.3	1486.2	867.4	594.0	547.4	375	1181	1144	375	246	2.00	2.30	3.30
MB	3228.1	3166.7	1835.6	1112.3	779.1	683.3	429	1247	1221	391	282	2.20	2.20	3.20
MG	1529.8	1498.1	926.5	607.2	457.3	413.8	209	535	525	196	85	2.40	2.20	3.50
Ml	3587.0	3523.2	2049.2	1250.6	880.8	773.3	424	1317	1272	415	264	2.20	2.30	3.50
ML	3254.8	3199.0	1881.4	1166.1	835.1	744.5	422	1271	1244	402	282	2.20	2.20	3.20
MP	1287.6	1260.7	731.9	427.0	293.8	266.7	178	470	466	154	88	2.30	2.20	3.20
MT	4564.6	4412.8	2317.0	1288.3	850.5	871.8	587	1862	1823	589	358	2.00	2.20	3.30
NG	3168.0	3115.1	1824.2	1107.5	771.0	663.6	406	1298	1270	413	285	2.10	2.20	3.20
NM	2905.4	2851.4	1612.5	958.2	662.7	577.5	381	1212	1181	380	271	2.20	2.20	3.20
OS	2289.5	2252.1	1309.9	784.8	544.9	483.8	292	763	751	238	178	2.10	2.30	3.50
PA	5751.4	5532.1	2915.7	1667.5	1105.3	1079.6	734	2453	2398	799	490	2.10	2.20	3.30
PF	2453.1	2409.0	1414.0	846.6	580.0	498.9	316	901	867	283	191	2.00	2.30	3.40
PG	3203.6	3146.9	1875.8	1154.1	813.9	711.5	424	1192	1156	374	254	2.20	2.20	3.30
PH	2491.1	2448.1	1448.4	875.7	604.3	526.1	323	914	882	288	196	2.00	2.20	3.40
PN	3211.2	3146.1	1784.3	1068.3	743.5	651.5	416	1331	1298	412	288	2.10	2.20	3.20
RC	5344.0	5163.4	2697.5	1526.3	1016.9	1098.9	670	2174	2122	711	427	2.10	2.20	3.40
RP	1466.1	1437.9	818.8	476.1	329.3	312.3	214	510	504	155	100	2.30	2.30	3.40
SC	4591.4	4455.4	2294.2	1285.9	859.3	835.6	561	1934	1889	596	402	2.00	2.20	3.30
ST	3071.1	2998.9	1650.3	984.3	683.9	636.4	403	1300	1277	404	280	2.10	2.20	3.10
TH	3624.3	3554.8	2018.7	1203.8	834.9	742.0	430	1374	1331	428	280	2.20	2.20	3.40
TM	2448.2	2396.7	1338.1	770.7	522.1	471.6	338	1004	976	302	223	2.10	2.20	3.20
TP	1331.5	1300.7	715.5	414.4	285.7	274.8	207	562	555	175	124	2.20	2.30	3.10

Table 2. Summary of some BINs included in the IWDB

BIN	Habitat Type	Location	#OA[a]	#OB[b]	BIN	Habitat Type	Location	#OAa	#OB[b]
1	Coral reefs	Pacific	10	26	21	Arctic community	Canada	29	86
2	Freshwater lake	Canada	7	29	22	Heathland habitat	Mauritius Is.	135	74
3	Freshwater lake	Canada	10	40	23	Beech forest	Japan	93	679
4	Freshwater lake	Canada	31	144	24	High Arctic	Canada	32	115
5	River	Canada	14	51	25	Montane forest	Australia	42	91
6	River	Canada	17	53	26	Multiple communities	Galapagos Is.	106	54
7	Freshwater lake	Canada	33	97	27	Xeric scrub	Argentina	21	45
8	Freshwater reservoir	Canada	6	25	28	Medow	UK	25	79
9	rainforest	Australia	51	41	29	Arctic community	Canada	II	18
10	rainforest	Peru	8	18	30	Deciduous forest	USA	13	44
11	tropical forest	C. Rica	6	4	31	Coastal forest	Mauritius Is.	14	13
12	Amazon rainforest	Brazil	16	25	32	Upland grassland	S. Africa	9	56
13	Arid grasslands	USA	54	24	33	Palm swamp	Venezuela	33	53
14	Whole country	Finland	5	64	34	Agricultural area	USA	456	1429
15	Andean scrub	Chile	87	98	35	Caatinga	Brazil	51	25
16	Boreal forest	Canada	12	102	36	Maple-oak woodland	USA	7	32
17	Caatinga	Brazil	13	13	37	Peat bog	Canada	13	34
18	MI. forest and grassland	USA	96	276	38	Montane forest	Argentina	10	29
19	High-altitude desert	Canary Is.	11	38	39	Forest	Papua	31	9
20	Alpine subarctic comm.	Sweden	23	118	40	Tropical forest	Panama	13	11

5.3. Data 3: V.S. Supreme Court Networks

We used a complex network constructed by Fowler *et al.* [4]. The authors included 26,681 majority opinions written by the U.S. Supreme Court. The dataset contains all cases that cite this U.S. Supreme Court decisions from 1791 to 2005. In this network each case is represented by a node. The links between two nodes a_i and a_j (arcs) express that the case j^{th} cites the i^{th} case previous to it (precedent). In order to carry out a study more focused on specific intervals of time and al so study the effect of accumulation of cites we split the data set of sub-sets (sub-networks). Each one of these sub- networks represent one slot of 5000 (5K) cases that cites 1000-5000 cases of the U.S. Supreme court in different decades. We call these networks as the 5K-Citations Network (5KCNs). It also makes the dataset more tractable for computation of $J_k(i)$ values. We constructed in total 43 5KCNs involving approximately 5000.43 > 22,000 cases of the U.S. Supreme court.

6. DISCUSSION

6.1. New Balaban-Markov Centralities for Nodes

As we mentioned in the introduction, Dehmer *et al. [43];* discussed this problem and suggested the classification of complex biological networks as a proof-of-concept experiment in this sense. They also concluded that there is a high necessity for freely available software packages to calculate and validate new TIs of complex networks. Different authors have reported, and/or used in QSAR studies, different types of TIs that may be considered variations or generalizations of the J index to some extent. For instance, Randić & Pompe [42]; introduced several variable molecular descriptors, derived from the distance matrix and the "reversed" distance matrix. This includes the variable Balaban J index and the "reversed" Balaban index 1/J as well as a novel index 1/JJ derived from 1 and 1/J. Balaban type indices: J, Jz, Jm, Jv, Jc, and Jp have been used to predict the supra-molecular complexing ability of a sulphonamides [47].

In this work we generalized the Balaban's J index in a different direction, implement it in a new software (MI-NODES), and carry out proof-of-concept experiments to demonstrate the discrimination power of the new indices. The product of the parameters $S_i \cdot S_j$ plays a central role in the definition of J (see eq. 1). In the introduction, S_i and S_j are distance sums calculated as the sums over the rows or columns of the topological distance matrix **D**. As a consequence, the result of a unique given graph G obtained only one value of J. However, we can weight this product with the probabilities k_{pij} with which both nodes are connected by walks of length k. In this case, we can generalize the J index to a series of higher order analogs J_k that quantify the probability of interconnection of all pairs of nodes at different orders. The values of kpij are the elements of the Markov matrix (1Π) used in the MARCH-INSIDE algorithm (see details in previous works). Also other authors, like Estrada [60], have used the same type of matrix to generate indices of complex networks more recently. In this work, we report for the first time this new type of J_k index. We also give the definition of the node centralities $J_k(i)$ for nodes of complex networks based on the same idea. In this case we can obtain a total of *k* values of new Balaban- Markov indices $J_k(G)$ and centralities $J_k(i)$ for each graph or each i^{th} node, using the following formula:

$$J_k(G) = \frac{q}{\mu + 1} \cdot \sum_{edges}^{q} \left({}^{k}p_{ij} \cdot S_i \cdot S_j \right)^{-\frac{1}{2}} \quad (6)$$

$$J_k(i) = \frac{q}{\mu + 1} \cdot \sum_{i-edges}^{z_i} \left({}^{k}p_{ij} \cdot S_i \cdot S_j \right)^{-\frac{1}{2}} \quad (7)$$

6.2. MIANN Models of MRNs

With the development of systems biology the study of Metabolic Reaction Networks (MRN s) is gaining importance due to possible applications in Biotechnology [61, 62] and Biomedicine with the study of disease comorbidity [63] which has been approached with network topology methods. In Table 1 we can find a summary of the properties of the MRNs studied. In this work, we used the $J_k(j)$ values and other parameters of MRNs presented in Table 1 to carry out a Two-Way Joining cluster analysis (TWJCA) of this dataset; see Fig. (2A).

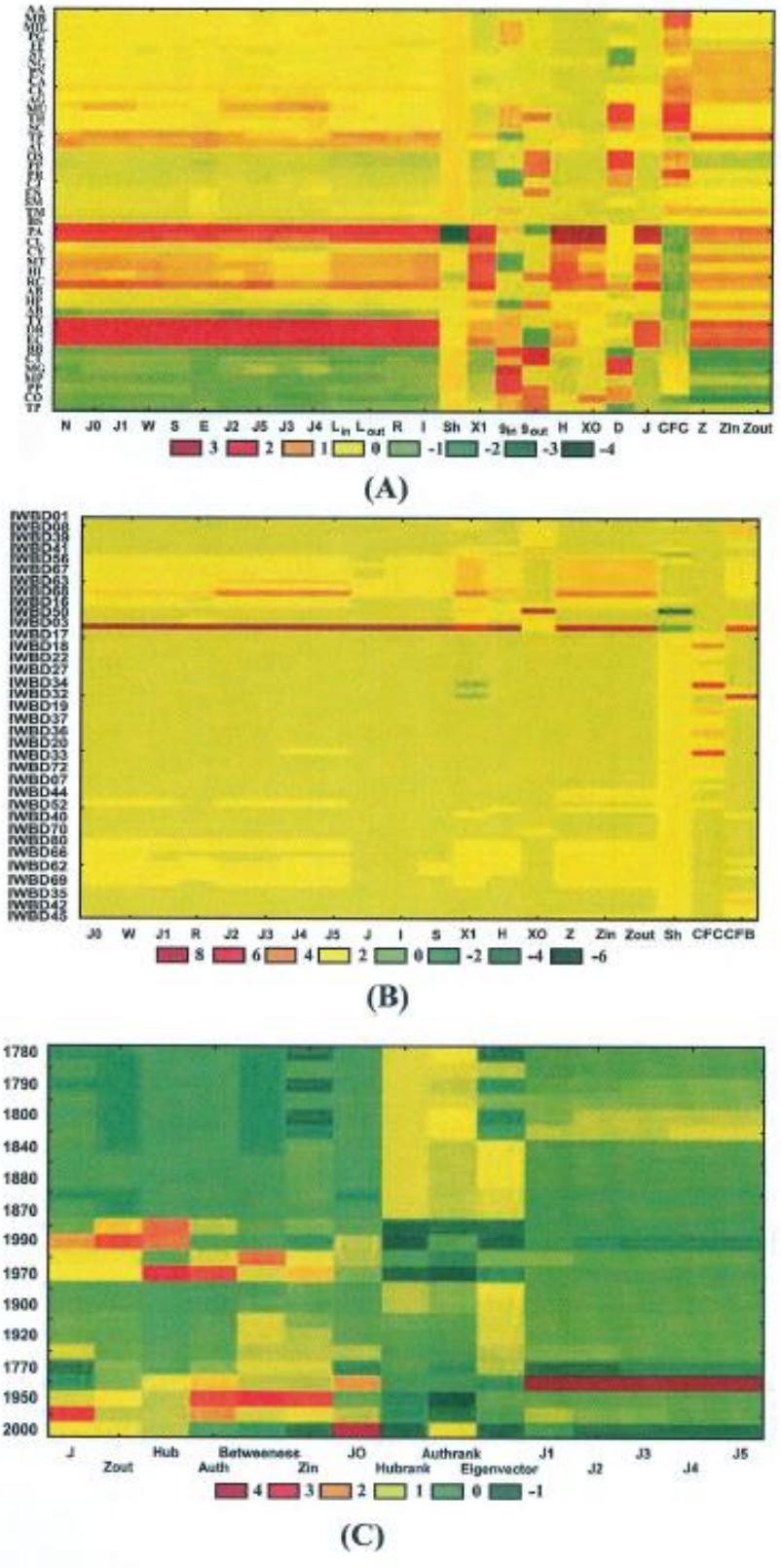


Fig. (2). Cluster analysis of $J_k(j)$ values vs classic TIs and other parameters of: (A) MRNs of 43 organisms, (8) BINs of 73 ecosystems, and (C) 73 US Supreme Court 5KCNs.

We used a Number of variables: 26 and Number of cases = 43 and found a Number of blocks = 193 with a Threshold computed from data = 0.4941518 (StDv/2), Total Sample Mean = -0.00, and Standard Deviation = 0.9883036. The results of this TWJCA are important to unravel hidden relationships between the new $J_k(j)$ indices with classic TIs as well as other parameters of the MRNs for each one of the 43 organisms studied. TWJCA shows that $J_k(j)$ indices form a "weak" cluster with some TIs for a group of networks but the TWJCA do not detect clear clusters. In general, TWJCA shows that $J_k(j)$ values seems to codify useful structural information of MRNs that it is not trivially related to the information codified by other parameters because there are not strong clusters formed between them.

Jeong *et al.* [2], noted that known MRNs models of different organisms present similar topological properties and not all metabolic pathways have been confirmed experimentally, and the experimental corroboration of a metabolome of one organism is a very hard task. Consequently, we need alignment-free techniques to evaluate the correct connectivity patterns L_i for all nodes in MRNs. Here we developed different MIANN models based on $J_k(i)$ values to predict correct connectivity patterns Li of nodes in MRNs of a large number of organisms (43 in total). The best MIANN model found was a non-linear three layers perceptron MLP 9:9-6-1:1. It has very good values of Accuracy (Ac), Sensitivity (Sn), and Specificity (Sp) > 81 % in both learning and external validation series. On the contrary, the best linear model did not show very good results with Ac, Sn, and Sp < 65% in all cases. This result indicates that $J_k(i)$ values are very powerful descriptors. Note that with a non-linear but relatively simple MLP (only 9 input s and 6 hidden neurons) we can predict in 43 organisms a very high number of 29117 metabolic reactions in learning and 9729 in validation. In Table 3 we depict the classification matrices and the topology of the MIANN models discussed.

6.3. MIANN Models of BINs

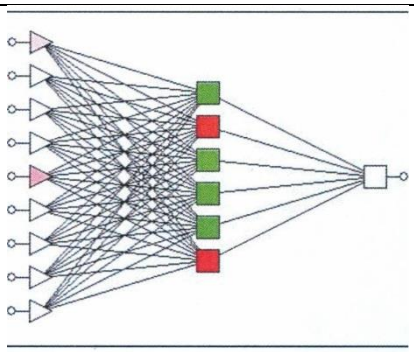
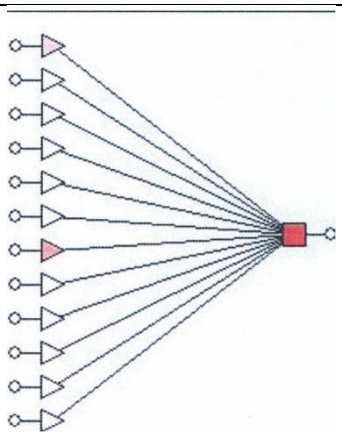
In analogy to the study of MRNs, we used the $J_k(j)$ values of BINs for different ecosystems and/or food webs. In Table 2 (se e previous section), we showed the names, location, number of species, and reference for many of these networks in order to illustrate the high complexity of the dataset studied. Almost all networks studied are directed and also bipartite in some cases. It means that the biological interaction is found in almost all cases between species of a group (A) with a given biological function with species of a second group with another function in many cases complementary to A somehow. For instance, we can find anemones, plants, or predators in the first group and parasite, herbivore, pollinator, prey, or seed disperser species in the second group. However, there are many complex networks with more complicated situations typical of an imbricate food web where many species may interact with the first or the second function in different cases. For instance, prey specie 1 is predated by predator 2, which is in turn predated / hunted by specie 3. Similarly, it may appear as hyperparasitism relationships where the host specie 1 is parasited by specie 2 which in tum is the host (is parasited by) of specie 3.

Here, we also used TWJCA in order to unravel relationships between $J_k(j)$ values (of BINs in this case) with other parameters of complex networks. In this second TWJCA experiment, we included classic TIs calculated with MI-NODES.

Some of these indices are the Wiener index (W), Shannon entropy (Sh), Gutrnan topological index (1), Schultz index (S) as well as χ connectivity indices of Randic or Kier & Hall. We also calculated with MI-NODES some node centralities of BINs like total, in, and out node degrees (Z, Z_{in}, Z_{out}) as well as Current Flow Centrality (CFC), and Current Flow Betweeness (CFB). TWJCA also pointed to $J_k(j)$ values as useful TIs that codify new structural information of BINs as they are not strongly clustered with other classic TIs. In Table 4 we depict the $J_k(j)$ values of many BINs. We also give Z_{in} and Z_{out} values for comparative purposes. We also added Wiener W values (proportional to the sum of all topological distance between the specie and all other species in the web. In Fig. (2B) we illustrate the results of the TWJCA experiment.

We also tested different MIANN models with linear and non-linear topology. Classification results in training and validation series for both MIANN models appear in Table 5. Linear MIANN models of BINs present slightly better results with respect to MRNs. The Ac, Sn, and Sp values here are close to 70% in all cases. However, once again the best model found is an MLP. The model MLP 12: 12-9-1: 1 present Ac, Sp, and Sn values higher than 81 %. The model is similar in goodness-of-fit and topology to the best MIANN model for MRNs.

Table 3. Linear *vs* Non-linear MIANN models of MRNs of 43 organisms based on J_k centralities.

MIANN Models	Li	Li=1	Li=0	%	Pr.	%	Li=1	Li=0
MLP 9:9-6-1:1	Li=1	29117	19490	81.8	Sn	81.9	9729	6426
	Li=0	6481	87304	81.7	Sp	81.9	2137	29172
LNN 12:12-1:1	Li=1	22320	39543	62.70	Sn	63.22	7502	13302
	Li=0	13278	67251	63.0	Sp	62.63	4364	22296

Pr. = Parameter, Sp = Specificity, Sn =Sensitivity. Columns: Observed classifications; Rows: Predicted classifications.

Table 4. Average values $J_k(i)_{org.avg}$ *vs* so me classic parameters of selected BINs of ecosystems.

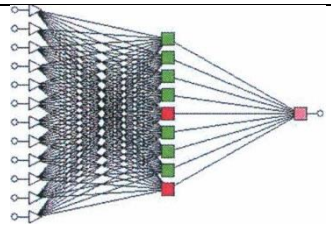
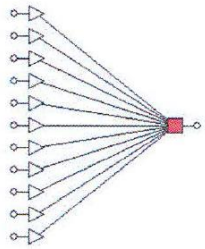
BIN	Name	J_o	J_1	J_2	J_3	J_4	J_5	z_{in}	z_{out}	W
IWBDOI	Anemone-fish	2.820	3.809	3.890	3.898	3.899	3.899	2.222	2.222	2.722
IWBD02	Aishihik lake	2.752	3.894	3.988	4.001	4.004	4.005	2.167	2.167	2.667
IWBD03	Cold lake	2.566	3.306	3.486	3.557	3.593	3.615	1.820	1.820	2.320
IWBD04	Lake of the woods	2.803	4.141	4.320	4.373	4.397	4.409	2.194	2.194	2.694
IWBD05	Mcgregor river	2.506	3.255	3.464	3.548	3.589	3.612	1.754	1.754	2.254
IWBD06	Parsnip river	2.918	3.747	3.925	3.986	4.013	4.027	2.257	2.257	2.757
IWBD07	Bay lake huron	3.028	4.162	4.274	4.298	4.308	4.314	2.431	2.431	2.931
IWBD08	Smallwood reservo ir	2.315	3.369	3.476	3.494	3.501	3.504	1.710	1.710	2.210
IWBD09	Bluthgen 2004	3.679	4.385	4.424	4.426	4.426	4.426	3.098	3.098	3.598
IWBDIO	Davidson *el al. 1989*	2.180	2.443	2.527	2.553	2.562	2.565	0.792	0.792	1.292
IWBDII	Davidson & Fisher	2.169	2.087	2.063	2.056	2.054	2.054	1.300	1.300	1.800
IWBDI2	Fonseca&ganade	2.066	2.287	2.353	2.377	2.388	2.395	l.l7l	l.l71	1.671
IWBDI3	Joern 1979 marathon	2.936	2.853	2.848	2.847	2.848	2.848	2.218	2.218	2.718
IWBDl4	J oern 1979 altuda	3.151	3.048	3.053	3.055	3.056	3.057	2.486	2.486	2.986
IWBDl5	Leather 1991 finland	2.031	5.848	6.313	6.391	6.411	6.419	1.377	1.377	1.877
IWBDI6	Leather 1991 britain	1.760	6.016	6.097	6.099	6.100	6.100	1.234	1.234	1.734
IWBDI7	Arroyo i	2.762	3.037	3.134	3.173	3.192	3.203	2.011	2.011	2.511
IWBD18	Arroyo ii	2.570	2.976	3.046	3.063	3.068	3.070	1.895	1.895	2.395
IWBD19	Arroyo iii	2.212	2.133	2.155	2.170	2.179	2.185	1.319	1.319	1.819
IWBD20	Barret & Helenurrn	2.228	4.124	4.789	5.108	5.278	5.374	1.465	1.465	1.965
IWBD21	Clements 1923	3.050	5.099	5.305	5.362	5.38a,	5.401	2.488	2.488	2.988
IWBD22	Dupont *el al. 2003*	2.847	3.750	3.923	3.984	4.013	4.031	2.163	2.163	2.663
IWBD23	Elberling & Olesen	2.352	3.422	3.693	3.815	3.879	3.917	1.688	1.688	2.188
IWBD24	Hocking 1968	2.193	3.904	4.109	4.186	4.226	4.251	1.600	1.600	2.100
IWBD25	Kato *el al. 1990*	2.336	5.292	6.113	6.415	6.550	6.621	1.566	1.566	2.066

6.4. MIANN Models of U.S. Supreme Court

There is a long tradition on the application of complex networks methods in social sciences; known as social network analysis (SNA) since 1930[64]. Using SNA we can unravel non-linear relationships between different laws and try to predict for instance the effect of these laws in society.

A turning point in this direction is the network constructed by Fowler *et al.* [4]. It represents a wonderful source (possibly the more complete) for the study of dynamics (changes along time) in the U.S. Supreme Court. The authors have withdrawn all cases cited in the text of each majority opinions from 1754 to 2002. According to them, opinion writers may cite a case just to mention it as a reference or because they disagree. Legal rules are cited to provide legal justifications even if it is not a reliance on authority. Thus, they included all judicial citations in the dataset (including various types of citations) that could link cases together.

Table 5. Linear *vs* non-linear MIANN models of BINs of 73 ecosystems.

MIANN Models	Li	Li=1	Li=0	%	Pr.	%	Li=1	Li=0
MLP 12:12-9-1:1	Li=1	3992	2684	81.6	Sn	81.3	1326	898
	Li=0	902	11934	81.6	Sp	81.6	305	3975
LNN 11:11-1:1	Li=1	3406	4435	69.6	Sn	69.3	1131	1470
	Li=0	1488	10183	69.7	Sp	69.8	500	3403

Pr. = Parameter, Sp = Specificity, Sn =Sensitivity. Columns: Observed classifications; Rows: Predicted classifications

SNA can be used to determine how important a case is to law at the Court and measure other legal concepts. However, the model is unable to predict the future evolution of these citations. In this type of situation, application of a model able to predict the future evolution of connectivity patterns $\mathbf{L}_i$ (direct and indirect citation patterns) of different cases along time may become a useful tool. For instance, our group has reported similar models for Spanish financial law network [52]. In this sense, it is straightforward to realise that here we should use TIs to describe the complex network data. The same authors, have used different node centralities to study this network before detecting more relevant cases (higher authority) at different times [5]. In addition, we should use a time-series technique if we want to predict the future evolution of case citation patterns. Last, as the problem is probably no-linear we should consider the probability of using a powerful non-linear algorithm to fit the data, as is the case of ANN s. All these features are present in MIANN models. Consequently, we decided to combine the new $J_k(i)$ centralities with MIANN analysis to model this data. First, we ca1culated the $J_k(i)$ values of all cases. Table 6 summarizes some of these values in two different time scales. In this table we give the average of $J_k(i)$ values for USSC in 5Ks cites slots *vs* decades of the citing case. Here 5K cites slot means that we construct a new sub-network for each period of time from the starting data in which appear 5000 more cites to U.S. Supreme Court cases. It means that 5K scale is divided into irregular periods of time but all sub-networks have the same number of links m = 5000 and different number of nodes (cases). Conversely, natural time scale networks contain different numbers of nodes (cases) and links (cites) accumulated in regular periods of time (decades).

Previously, to carry out the MIANN analysis we decided to explore the possible relationships between the $J_k(i)$ values and other node centralities previously used to describe this network. We carried out a TWJCA of this dataset as well; see Fig. (2C). We used a Number of variables = 15 and Number of cases = 25 decades. The other variables used, in addition to $J_k(i)$ values, were the following: Z_{in} = in degree, Z_{out} = out degree, Hub = raw Hub score, Hubrank = rank of hub score, Auth = raw Authority score, Authrank = Rank of Authority score, Betweeness = 5KCN Node betweeness centrality for a give case, Eigenvector = eigenvector centrality measure.

We found a Number of blocks = 112 with a Threshold computed from data = 0.4898979 (StDv/2), Mean = -0.00, and Standard Deviation = 0.9797958. TWJCA shows that $J_k(j)$ indices of the U.S. Supreme Court with degree k = 1 to 5 form their own cluster. The effect is stronger in the decades from 1840 to 1990. It is interesting that $J_0(G)$ does not form clusters with the other $J_k(j)$ indices. In general $J_k(j)$ indices do not form clusters with other TIs or no de centralities of the U.S. Supreme Court. The MIANN model here was also the MLP, now with slightly better goodness-of-fit with respect (Ac, Sn, and Sp approximately 82-83%) to MRNs and BINs (Ac, Sn, and Sp approximately 81%). Interestingly, the best linear model is notably better here with (Ac, Sn, and Sp approximately 79%), which is lower but similar to the MLP. Considering that both models MLP 18:18-10-1:1 and LNN 18:18-1:1 has very similar performance and the same number of inputs we should consider the simpler LNN (18 variables and O hidden neurons) model also as a very good model. The MLP needs 10 hidden neurons to increase performance only in 1-2%. The situation is curious, linear models were increasing in performance from bio-molecular process to ecological and social systems.

CONCLUSION

In this work, we report for the frist time a new class of Balaban type parameters called Balaban-Markov centralities. Contrary to tradition, the new indices were not defined for small molecules but for all classes of systems susceptible to be represented for graphs. We also report three proof-of- concept experiments, to test the power of the new $J_k(i)$ indices to predict actual no de connectivity patterns in complex bio-molecular, ecological, and social networks.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The authors acknowledge the kind attention of the Editor-In-Chief Prof. Alessandro Giuliani Istituto Superiore di Sanitá (Italian NIH) Environment and Health Dept Rome, Italy.

REFERENCES

[1] Boccaletti S, Latora V, Moreno Y, *et al.* Complex networks: Structure and dynamics, Physics Reports 2006; 424: 175-308.

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

[3] Estrada E. Food webs robustness to biodiversity loss: the roles of connectance, expansibility and degree distribution. J Theor Biol 2007; 244: 296-307.

[4] Fowler JH, Jeon S. The authority of Supreme Court precedent, Social Networks 2008; 30: 16-30.

[5] 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.

[6] Estrada E. Returnability as a criterion of disequilibrium in atmospheric reactions network. J Math Chem 2012; 50: l363-72.

[7] Rosenquist JN, Murabito J, Fowler JH, *et al.* The spread of alcohol consumption behavior in a large social network. Ann Intern Med 2010; 152: 426-33.

[8] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007; 357: 370-9.

[9] Bornholdt S, Schuster HG. Handbook of Graphs and Complex Networks: From the Genome to the Internet. Wheinheirn: WILEY- VCH GrnbH & CO. KGa., 2003.

[10] Sporns O, Tononi G. Classes of network connectivity and dynamics. Complexity 2002; 7: 28-38.

[11] Estrada E, Uriarte E. Recent advances on the role of topological indices in drug discovery research. Curr Med Chem 2001; 8: 1573- 88.
[12] García-Domenech R, Gálvez J, de Julian-Ortiz JV, *et al.* Some new trends in chemical graph theory. Chem Rev 2008; 108: 1127-69.
[13] Komosinski M, Kubiak M. Quantitative measure of structural and geometric similarity of 3D morphologies. Complexity 2011; 16(6): 40-52.
[14] Diudea MV, Aleksander I, Kurt V, *et al.* Network analysis using a novel highly discriminating topological index. Complexity 2011; 16(6): 32-9.
[15] Dehmer M, Department for Biomedical Informatics and Mechatronics ItBaTR, UMIT, Eduard Wallnoefer Zentrum 1, A- 6060, Hall in Tyrol, Austria, Department for Biomedical Informatics and Mechatronics ItBaTR, UMIT, Eduard Wallnoefer Zentrum 1, A-6060, Hall in Tyrol, Austria *et al.* Generalized graph entropies. Complexity 2011; 17(2): 45-50.
[16] T BA. HIGHLY DISCRIMINATING DISTANCE-BASED TOPOLOGICAL INDEX. Chem Phys Lett 1981; 89: 399-404.
[17] Basak SC, Balaban AT, Grunwald GD, *et al.* Topological indices: their nature and rnutual relatedness. J Chern Inf Comput Sci 2000; 40: 891-8.
[18] 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.
[19] 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.
[20] Wiener H. Vapor pressure-temperature relationships among the branched paraffin hydrocarbons. J Phys Colloid Chem 1948;52:425-30.
[21] 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.
[22] Hosoya H. Mathematical meaning and importance of the topological index Z. Croat Chem Acta 2007;80:239-49.
[23] Randic M. Characterization of molecular branching. J Am Chem Soc 1975;97:6609-15.
[24] Kier LB, Hall LH, Murray WJ, *et al.* Molecular connectivity. I: Relationship to nonspecific local anesthesia. J Pharm Sci 1975; 64: 1971-4.
[25] Hall LH, Kier LB. Issues in representation of molecular structure the development of molecular. J Mol Graph Model 2001; 20: 4-18.
[26] Balaban AT. From chemical graphs to 3D molecular modeling. In: Balaban A. T. (ed) From chemical topology to three-dimensional geometry. New York: Plenum Publishers, 1997,420.
[27] Sharma SK, Kumar P, Narasirnhan B, *et al.* Synthesis, antimicrobial, anticancer evaluation and QSAR studies of 6- methyl-4-[1-(2-substituted-phenylamino-acetyl)-1 H -indol-3-yl]-2- oxo/thioxo-I,2,3,4-tetrahydropyrirnidine-5-carboxylic acid ethyl esters. Eur J Med Chem 2012; 48: 16-25.
[28] Thakur A, Thakur M, Khadikar PV, *et al.* QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: topological. Bioorg Med Chem 2004; 12: 789-93.
[29] Krawczuk A, Voelkel A, Lulek J, *el al.* Use of topological indices of polychlorinated biphenyls in structure-retention. J Chromatogr A 2003; 1018: 63-71.
[30] 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.
[31] Naik PK, Dubey A, Kumar R. Development of predictive quantitative structure-activity relationship models of. J Biomol Screen 2010; 15: 1194-203.
[32] Fernandes JP, Pasqualoto KF, Felli VM, *et al.* QSAR modeling of a set of pyrazinoate esters as antituberculosis prodrugs. Arch Pharm (Weinheim) 2010; 343: 91-7.
[33] 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.
[34] Zavrsnik D, Muratovic S, Spirtovic S. QSAR and QSPR study of derivatives 4-arylaminocoumarin. Bosn J Basic Med Sci 2003; 3: 59-63.
[35] Ma XL, Chen C, Yang J. Predictive model of blood-brain barrier penetration of organic compounds. Acta Pharmacol Sin 2005; 26: 500-12.
[36] Dashtbozorgi Z, Golmohammadi H. Quantitative structure-property relationship modeling of water-to-wet butyl. J Sep Sci 2010; 33: 3800-10.
[37] Jalali-Heravi M, Fatemi MH. Prediction of thermal conductivity detection response factors using an artificial neural network. J Chromatogr A 2000; 897: 227-35.
[38] Dehmer M, Grabner M, Varmuza K. Information indices with high discriminative power for graphs. PLoS One 2012;7:e31214.
[39] Basak SC, Mills D, Gute BD, *et al.* Use of mathematical structural invariants in analyzing combinatorial libraries: a case study with psoralen derivatives. Curr Comput Aided Drug Des 2010; 6: 240- 51.
[40] Ratkiewicz A, Truong TN. Application of chemical graph theory for automated mechanism generation. J Chem Inf Comput Sci 2003; 43:36-44.

[41] Rastija V, Medic-Saric M. QSAR study of antioxidant activity of wine polyphenols. Eur J Med Chem 2009;44:400-8.
[42] Randic M, Pompe M. The variable molecular descriptors based on distance related matrices. J Chem InfComput Sci 2001;41:575-81.
[43] Mueller LA, Kugler KG, Graber A, *et al.* Structural Measures for Network Biology Using QuACN. BMC Bioinformatics 2011;12:492.
[44] Junker BH, Koschutzki O, Schreiber F. Exploration of biological network centralities with CentiBiN. BMC Bioinformatics 2006; 7: 219.
[45] Batagelj Y, Mrvar A. Pajek- Analysis and Visualization of Large Networks. Lect Notes Comput Se 2002; 2265: 477-8.
[46] Mueller LA, Kugler KG, Dander A, *et al.* QuACN: an R package for analyzing complex biological networks quantitatively. Bioinformatics 2011;27:140-l.
[47] Balaban A T, Khadikar PY, Supuran CT, *et al.* Study on supramolecular complexing ability vis-a-vis estimation of pKa of substituted sulfonamides: Dominating role of Balaban index (J). Bioorg Med Chem Lett 2005;15: 3966-73.
[48] 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.
[49] 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.
[50] Riera-Fernández P, Munteanu CR, Martin-Romalde R *et al.* Markov-Randic Indices for QSPR Re-Evaluation of Metabolic, Parasite-Host, Fasciolosis Spreading, Brain Cortex and Legal- Social Complex Networks. Curr Bioinformatics 2013;8:401-15.
[51] Gonzalez-Díaz H, Riera-Fernández 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: 3331-40.
[52] 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.
[53] Hill T, Lewicki P. STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining. Tulsa: StatSoft 2006.
[54] Frank E, Hall M, Trigg L, *et al.* Data mining in bioinforrnatics using Weka. Bioinformatics. England, 2004; 2479-8l.
[55] González-Díaz H, Arrasate S, Sotomayor N, *et al.* MIANN Models in Medicinal, Physical and Organic Chemistry. Curr Top Med Chem 2013;13: 619-4l.
[56] Box GEP, Jenkins GM. Time series analysis. Holden-Oay, 1970.
[57] Hill T, Lewicki P. STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining. Tulsa: Stat Soft, 2006.
[58] Riera-Fernández P, Munteanu CR, Pedreira-Souto N, *et al.* Definition of Markov-Harary Invariants and Review of Classic Topological lndices and Databases in Biology, Parasitology, Technology, and Social-Legal Networks. Curr Bioinformatics 2011;6:94-12l.
[59] 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 Bioinformatics 2011;6:94-12l.
[60] Estrada E. Information mobility in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 2009; 80: 026104.
[61] Mazurie A, Bonchev O, Schwikowski B, *et al.* Evolution of metabolic network organization. BMC Syst Bio12010; 4: 59.
[62] Kier LB, Bonchev O, Buck GA. Modeling biochemical networks: a cellular-automata approach. Chem Biodivers 2005; 2: 233-43.
[63] Lee OS, Park J, Kay KA, *et al.* The implications of human metabolic network topology for disease comorbidity. Proceedings of the National Academy of Sciences of the United States of America 2008; 105:9880-5.
[64] Wasserman S, Faust K. Social network analysis: methods and applications. Cambridge: Cambridge University Press, 1999.