

Biosystems. 2013; 111(3): 199-207

The Rucker–Markov invariants of complex bio-systems: applications in parasitology and neuroinformatics

Humberto González-Díaz^{a,b}, Pablo Riera-Fernández^c, Alejandro Pazos^d, Cristian R. Munteanu^d

^a *Department of Organic Chemistry II, Faculty of Science and Technology, Euskal Heriko Unibertsitatea (UPV/EHU), Leioa, 48940, Bizkaia, Basque Country, Spain*

^b *IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

^c *Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain*

^d *Computer Science Faculty, University of A Coruña, 15071 A Coruña, Spain*

Abstract

Rucker's walk count (WC) indices are well-known topological indices (TIs) used in Chemoinformatics to quantify the molecular structure of drugs represented by a graph in Quantitative structure–activity/property relationship (QSAR/QSPR) studies. In this work, we introduce for the first time the higher-order (k th order) analogues (WC_k) of these indices using Markov chains. In addition, we report new QSPR models for large complex networks of different Bio-Systems useful in Parasitology and Neuroinformatics. The new type of QSPR models can be used for model checking to calculate numerical scores $S(L_{ij})$ for links L_{ij} (checking or re-evaluation of network connectivity) in large networks of all these fields. The method may be summarized as follows: (i) first, the $WC_k(j)$ values are calculated for all j th nodes in a complex network already created; (ii) A linear discriminant analysis (LDA) is used to seek a linear equation that discriminates connected or linked ($L_{ij} = 1$) pairs of nodes experimentally confirmed from non-linked ones ($L_{ij} = 0$); (iii) The new model is validated with external series of pairs of nodes; (iv) The equation obtained is used to re-evaluate the connectivity quality of the network, connecting/disconnecting nodes based on the quality scores calculated with the new connectivity function. The linear QSPR models obtained yielded the following results in terms of overall test accuracy for re-construction of complex networks of different Bio-Systems: parasite–host networks (93.14%), NW Spain fasciolosis spreading networks (71.42/70.18%) and CoCoMac Brain Cortex co-activation network (86.40%). Thus, this work can contribute to the computational re-evaluation or model checking of connectivity (collation) in complex systems of any science field.

Keywords

Complex networks; Parasite–host networks; Brain Cortex network; Walk count; Markov chains; Graph topological indices; Quantitative structure–property relationships

1. Introduction

The study of networks lies within the frontiers of many science branches, from Chemistry and Molecular Biology to Neurobiology, Internet, World Trade, or social sciences. The most basic issues are structural: how does one characterize the connectivity patterns in those networks? Are there any unifying features underlying their topology? Different research groups have begun to shed light over these unifying aspects of the structure and dynamics of complex networks (Boccaletti et al., 2006; Bornholdt and Schuster, 2003; Dehmer and Emmert-Streib, 2009; Newman et al., 2002; Ratti et al., 2010; Strogatz, 2001, 2005; Watts and Strogatz, 1998). Networks are represented by means of a graph as a way to capture essential information. Graphs in turn are sets of items, drawn as dots, or *nodes*, interconnected by lines or arcs, which represent wires, ties, links, edges, bonds, or pair-wise relationships in general. Consequently, the nodes can represent atoms, molecules, proteins, nucleic acids, drugs, cells, organisms, parasites, people, words, laws, computers or any other part of a real system. Moreover, the lines represent relationships between the nodes such as chemical bonds, physical interactions, metabolic pathways, pharmacological actions, law recurrence or social ties (Bonchev, 2003, 2004; Bonchev and Buck, 2007; Bonchev and Rouvray, 2005; Duardo-Sanchez et al., 2011; Gonzalez-Diaz, 2010a,b; Newman, 2003; Thomas and Bonchev, 2010; Vina et al., 2009).

There are different experimental and/or theoretical methods that assign node–node links depending on the type of network to be created. 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 Quantitative Structure–Property/Activity Relationship (QSPR/QSAR) models. This methodology was traditionally used in Chemoinformatics. QSARs are used to predict the biological activity of drugs whereas QSPRs are used to model physicochemical properties of organic compounds. QSPR-like models use as input mostly structural parameters derived from the graph representation of the network-like system under study (Puzyn et al., 2010). Many authors refer to the numerical parameters of a graph as topological indices (TIs), mainly in the case of global studies (properties of full system). Node centralities or local TIs of a sub-graph can be used if we want to predict a local property of a given part of the system. Currently, the use of QSPR-like models in which the inputs are graph parameters is not limited to the study of molecules and has been extended to other complex systems (González-Díaz and Munteanu, 2010).

There are many classes of TIs useful to quantify the structure of a system in QSPR-like studies. In this sense, Rucker and Rucker (2000) published a work about the use of walk count (WC) indices. In this previous work, it was demonstrated how the complexity of a (molecular) graph can be quantified in terms of walk counts, obtaining extremely easy graph invariants that depend on size, branching, cyclicity, and edge and vertex weights (w). Weights are important to numerically quantify properties that differentiate classes of nodes. This is the case of unsaturation or heteroatoms in small molecules. In 2001, the same authors (Rucker and Rucker, 2001) returned to the idea of WCs with a discussion about unsaturated compounds represented by multi-graphs and the necessity to distinguish between the notions of sub-structure and sub-graph. They introduced a new computer program able to construct and count all sub-structures and sub-graphs for a colored multi-graph (*e.g.*, a molecular compound which may contain unsaturations and heteroatoms).

In the same year, Gutman co-authored with Rucker and Rucker another paper about WCs (Gutman et al., 2001). They highlighted that WCs have been used for a long time in applications of Theoretical Chemistry based on the fact that the (i, j) -entry of the k th power of the adjacency matrix is equal to the number of walks starting at vertex i , ending at vertex j , and having length k . They discussed about the numbers of all walks of length k . These values are called molecular walk counts, $mwc(k)$, in the case of molecular graphs. They also commented that the sum of $mwc(k)$ from $k = 1$ to $k = n - 1$, called total walk count (twc), as quantities that are suitable for QSPR studies and capable of measuring the complexity of organic molecules. Next, they established a few general properties of mwc 's and twc among which there are the linear dependence between the mwc 's, the spectral decomposition of mwc 's, and various connections between the walk counts and the eigenvalues and eigenvectors of the molecular graph. In 2003, the concept of atomic wc and

mwc was expanded by Lukovits and Trinajstić (2003) to zero and negative orders. They used a backward algorithm based on the usual procedure employed to obtain the values of mwc's. More recently, Bonchev has applied WCs and other TIs to the complexity analysis of yeast proteome network (Bonchev, 2004).

All the above-mentioned aspects paved the way for at least two directions in the research of WCs. First, they focus on general walk counts of length k (WC_k) as these TIs are promising for the QSPR study of any class of network-like systems, not only small molecules. In addition, the extension of WCs calculations using different mathematical models, seems to be suitable. Our group has introduced the software called MARCH-INSIDE (Markovian Chemicals In Silico Design), or shortly MI, which has become a very useful tool for QSAR/QSPR studies for drugs, proteins, and more complex systems (Aguero-Chapin et al., 2008a,b, 2009; Concu et al., 2009; Gonzalez-Díaz et al., 2008, 2010a,b,c, 2011; González-Díaz et al., 2003, 2007; Munteanu et al., 2009; Ramos de Armas et al., 2004; Rodríguez-Soca et al., 2010; Santana et al., 2008). However, MI can perform a limited management of other complex networks. Recently, we have re-programmed the MI application creating a new software application able to manage complex networks. The new program is called MI-NODES (MARCH-INSIDE Node DEScriptors) and is compatible with other software such as Pajek or CentiBin (it is able to upload .mat, .net and .dat formats). A very interesting feature of MI-NODES is that it can calculate general versions of classic molecular TIs for large complex networks using Markov Chains theory. In previous works, we have introduced several types of Markov TIs such as Markov–Shannon Entropy (Berca et al., 2011), Markov–Randić indices (Aguar-Pulido et al., 2010), or Markov–Harary numbers (Riera-Fernández et al., 2011). In these works, we have used the Markov-TIs in order to compare several types of complex networks from different fields such as Biology Linguistics, Technology, Sociology and Law. The obtained results have shown the usefulness of the Markov-TIs in network-based studies.

In this work, we introduce the new Rucker–Markov indices and use them, for the first time, as inputs to seek QSPR-like models able to assess the quality of the connectivity of new complex networks. The idea is to search for a QSPR-like model that uses as input the $WC_i(j)$ values for all possible pairs of nodes in a network to decide which pairs of nodes are linked and which are not. This class of model will allow us to computationally re-evaluate all the links 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. In order to illustrate the use of the new method, we have used three types of networks. In each case, we report for the first time new QSPR models useful to re-evaluate connectivity quality of different types of networks. The flowchart of the method is shown in Fig. 1. In the first case, we studied different biological networks of parasite–host interactions (PHIs). In the second case, we used two macroscopic landscape parasitism–spreading networks for cattle fasciolosis in NW Spain. The third case consisted of carrying out a study regarding connectivity quality in the CoCoMac cerebral cortex co-activation network (Modha and Singh, 2010). Therefore, this method will bring new possibilities for the assessment of the connectivity quality in new complex networks.

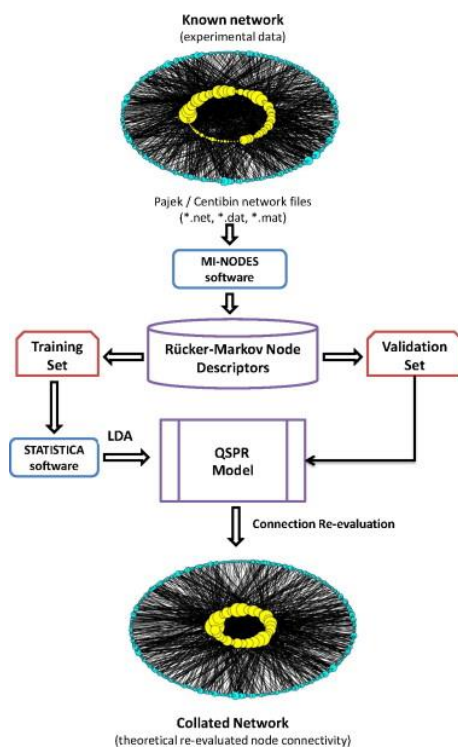


Fig. 1. General workflow for connectivity re-evaluation in complex networks.

2. Materials and methods

2.1. Datasets used

2.1.1. Parasite–host complex networks

In this work we have used four bipartite parasite–host networks: parasite–fish, parasite–ungulates, parasite–carnivores and parasite–primates, which have been reported before by our group (Riera-Fernandez et al., 2012). In these networks the first set of nodes represent the parasites and the second one the hosts. Two nodes are connected if the parasite interacts with the host.

2.1.2. Complex network for fasciolosis spreading in NW Spain

The nodes of these networks are represented by farms. The Boolean or connectivity matrix \mathbf{C} quantifies the propensity $C_{ij} = 1$ of disease to spread between farms immediately after treatment. On the other hand, matrix \mathbf{L} includes two criteria: the pre-existence of a high propensity of disease spreading $C_{ij} = 1$ and the experimental confirmation of a high risk ratio (RR_{ij}) of prevalence after treatment (PAT_j) of the disease. See the definition of these networks in mathematical terms in our previous works (Riera-Fernandez et al., 2012). The original dataset used here was collected by Mezo et al. (2008) and employed to create a network of farm-to-farm spreading of fasciolosis in cattle in Galicia (NW Spain), discussed in other work carried out by our group (González-Díaz et al., 2010).

2.1.3. Macaque cerebral cortex co-activation network

We use the information contained in the Collation of Connectivity data on the Macaque (CoCoMac) brain network (Modha and Singh, 2010), a Neuroinformatics dataset consisting of 383 hierarchically organized regions spanning cortex, thalamus, and basal ganglia, with 6602 directed long-distance connections (<http://www.cocomac.org>) (see Fig. 2).

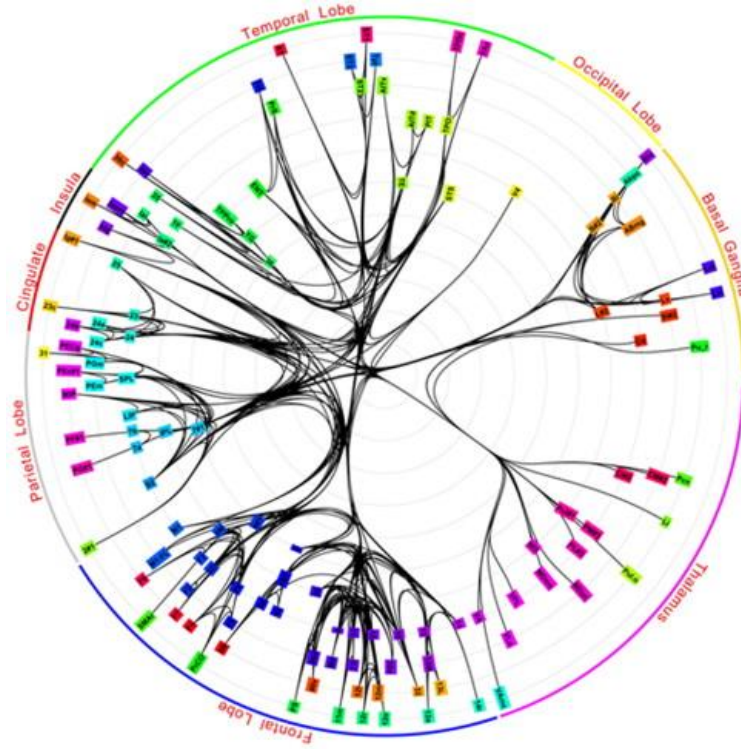


Fig. 2. Innermost core for the undirected version of CoCoMac network. The innermost core is a central sub-network that is far more tightly integrated than the overall network.

2.2. Computational methods

2.2.1. The Markov–Rücker centralities

We implemented the new centralities in the MI-NODES (MARCH-INSIDE for Node DEscriptors) software and used them to calculate the node centralities of the networks studied in this work. MI-NODES is a GUI Python/wxPython application developed by our group that can manage any kind of complex network. The program builds a Markov matrix (${}^1\Pi$) for each network using as input the connectivity matrix L . The elements of this stochastic matrix are the node–node transition probabilities (p_{ij}). The probability matrix is raised to the power k , resulting $({}^1\Pi)^k$. The resulting matrices ${}^k\Pi$, which are the k th natural powers of ${}^1\Pi$, contain the transition probabilities ${}^k p_{ij}$. 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–Rücker indices of order k th is straightforward to carry out by simply substituting the values of d_{ij} with these probabilities ${}^k p_{ij}$. In so doing, we can obtain k values of the new Markov–Rücker (or probabilistic walk counts) indices $WC_k(G)$ for a graph G . 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 is equal to δ_i (the degree of i)). In this simple case we can obtain a total of k values of new Markov–Rücker or probabilistic walk count centralities $WC_k(i)$ for the node i th; see equations:

$$WC_k(G) = \frac{1}{2} \sum_{i=1}^D \sum_{j=i}^D {}^k p_{ij} \quad (1)$$

$$WC_k(j) = \frac{1}{2} \sum_{i=1}^1 \sum_{j \rightarrow i}^{\delta_i} {}^k p_{ij} = \frac{1}{2} \sum_{j \rightarrow i}^{\delta_i} {}^k p_{ij} \quad (2)$$

2.2.2. MI-NODES software for the calculation of Markov indices

MI-NODES calculates new types of node Centralities ${}^k C_c(j)$ 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 can read Pajek and CentiBin networks such as *.net, *.dat and *.mat.

2.2.3. Linear discriminant analysis (LDA) models

Linear discriminant analysis (LDA) is possibly the most common technique used in QSPR/QSAR studies with TIs of molecular graphs, protein and RNA structure networks, and bio-molecular complex networks. Let be $S(L_{ij})$ the output variable of a model used to score the quality of the connection between two nodes i th and j th ($L_{ij} = 1$). We can use LDA to seek a linear equation with coefficients a_i , a_j , a_{ij} and a_0 (Eq. (3)). These are the coefficients of the TIs used as input (in this case local node centralities) in the QSPR-like model and the independent term. The terms a_{ik} and a_{jk} refer to all nodes that lie within the k th neighborhood (placed at least at a topological distance $d = k$) of i th or j th single nodes respectively. The term a_{ijk} refers to the differences between the neighborhoods of a pair of nodes, which may be connected or not. We can use different statistical parameters to evaluate the statistical significance and validate the goodness-of-fit of LDA equation: n = number of cases, χ^2 = Chi-square, p = error level, as well as accuracy, specificity, and sensitivity of both training and external validation series (Hill and Lewicki, 2006). We can write a general linear equation for an LDA-based QSPR/QSAR model with the parameters mentioned above. In general, this type of LDA-based QSPR/QSAR models has been used before for many research studies, mainly for small molecules (Casanola-Martin et al., 2006, 2007, 2008; Estrada et al., 2001, 2002, 2006, 2010; Marrero-Ponce et al., 2007a,b, 2010; Speck-Planche and Cordeiro, 2011; Speck-Planche et al., 2009, 2011a,b,c; Vilar et al., 2005). The LDA model based on the new Rucker–Markov indices is shown in the following formula:

$$S(L_{ij}) = \sum_{k=0}^5 a_{ik} \cdot WC_k(i) + \sum_{k=0}^5 a_{jk} \cdot WC_k(j) + \sum_{k=0}^5 a_{ijk} \cdot [WC_k(i) - WC_k(j)] + a_0 \quad (3)$$

2.2.4. Graphical representation and description of the networks

The information used to create the networks was obtained from connectivity matrices in the case of the observed networks and from the output of the LDA models in the case of the reconstructed networks. CentiBin (v.1.4.3) (Junker et al., 2006) was used to prepare the networks in *.mat files (<http://centibin.ipk-gatersleben.de/index.php>) in order to calculate the graph diameter, Rucker index and average distance. The density (taking into account whether the graph is unipartite or bipartite), average degree and Randić index were also calculated using Pajek (v.1.26) (Batagelj and Mrvar, 1998; De Nooy et al., 2005). This program was used to represent graphically the prepared versions of the observed and reconstructed networks (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>).

3. Results and discussion

3.1. Model 1: parasite–host networks

The parasite–host interactions (PHIs) became increasingly important to understand the role of phylogenetic and ecological factors on the parasite–host specificity (Desdevises et al., 2002; Detwiler and Janovy, 2008; Poulin et al., 2011) and to know how parasites affect the ecosystem functioning (Anderson and May, 1979; Hatcher et al., 2006; Price et al., 1986). Thus, network theory became a useful tool for analyzing this type of interaction (Poulin, 2010). Due to the high experimental difficulty inherent to the *in situ* accurate determination of PHIs, the computational models are very effective tools to correct the PHIs networks. In this work, we have used χ_k to seek a QSPR-like model able to score the quality of PHIs in known networks. The best model found was the following:

$$\begin{aligned} S(L_{ij}) = & -258.93 \cdot [WC_1(p_i) - WC_1(h_j)] + 283.69 \cdot \\ & [WC_2(p_i) - WC_2(h_j)] - 88.75 \cdot [WC_4(p_i) - WC_4(h_j)] + 0.25 \end{aligned} \quad (4)$$

$n = 49,218; \chi^2 = 22,297; p < 0.001$

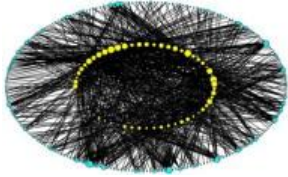
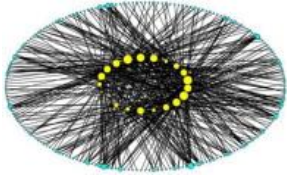
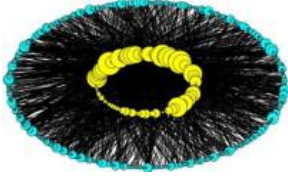
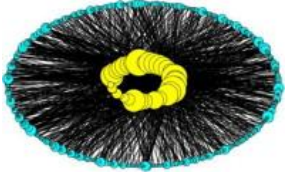
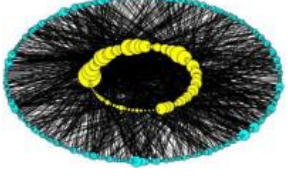
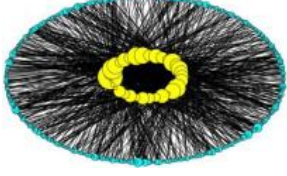
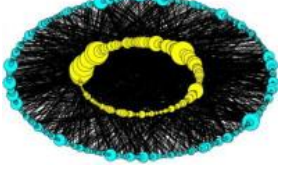

In this equation, $S(L_{ij})$ is a real-valued output variable that scores the propensity of the i th parasite specie (p_i) to infect a given host specie (h_j). The Chi-square statistic (χ^2) shows a low p -level < 0.001 , which indicates a significant discrimination between well-established host–parasite relationships and not confirmed parasitism. The model shows good values of accuracy, sensitivity, and specificity for the recognition of parasite–host relationships (links) both in training and external validation series (see Table 1, Model 1). Consequently, with this simple linear model we could re-evaluate connectivity quality in the already known PHIs networks in a fast and non-expensive way (without experimentally resampling all PHIs in the corresponding ecological niche). The giant components of the observed and reconstructed networks are described numerically and graphically in Table 2.

Table 1. Training and validation results for all models developed in this work.

QSPR model	Training series			Model parameters	Validation series		
	NL	L	%		%	NL	L
1	Parasite–host networks						
	42,503	2125	NL 95.24	Specificity	95.43	NL 14,120	676
	1227	3363	L 73.27	Sensitivity	70.81	L 444	1077
			Total 93.19	Accuracy	93.14	Total	
2a	NW Spain fasciolosis landscape-spreading network (C)						
	15,000	5983	NL 71.49	Specificity	71.55	NL 5035	2002
	902	2278	L 71.64	Sensitivity	70.54	L 312	747
			Total 71.51	Accuracy	71.42	Total	
2b	NW Spain fasciolosis landscape-prevalence network (L)						
	15,934	6688	NL 70.44	Specificity	70.16	NL 5337	2270
	404	965	L 70.49	Sensitivity	70.44	L 133	317
			Total 70.44	Accuracy	70.18	Total	
3	Cerebral cortex co-activation network						
	30,161	3957	NL 88.40	Specificity	88.30	NL 10,068	1334
	1256	3696	L 74.64	Sensitivity	73.27	L 441	1209
			Total 86.66	Accuracy	86.40	Total	

Rows: observed classifications; columns: predicted classifications.

Table 2. Comparison of original vs. re-constructed parasite–host networks (giant components).

Observed network	Network descriptors ^a			Reconstructed network
Parasites–fish				
	298	n	229	
	239	np	208	
	59	nh	21	
	912	m	505	
	6.12	Ad	4.41	
	0.0647	den	0.116	
	95.77	R	58.40	
	298,626	W	178,232	
	7	D	6	
	3.37	AD	3.41	
Parasites–ungulates				
	793	n	716	
	701	np	677	
	92	nh	39	
	1863	m	1439	
	4.70	Ad	4.02	
	0.0289	den	0.0545	
	191.02	R	142.48	
	2,534,140	W	1,971,820	
	10	D	8	
	4.03	AD	3.97	
Parasites–carnivores				
	619	n	542	
	537	np	508	
	82	nh	34	
	1343	m	1014	
	4.34	Ad	3.74	
	0.0305	den	0.0587	
	159.16	R	117.23	
	1,587,048	W	1,181,852	
	9	D	9	
	4.15	AD	4.14	
Parasites–primates				
	913	n	766	
	757	np	704	
	156	nh	62	
	1993	m	1482	
	4.37	Ad	3.87	
	0.0169	den	0.0340	
	252.66	R	180.73	
	3,609,008	W	2,469,592	
	10	D	12	
	4.33	AD	4.27	

^a Network descriptors: total number of connected nodes (n), number of connected parasites (np), number of connected hosts (nh), number of edges (m), average degree (Ad), density (den), Randić connectivity index (R), Rucker index (W), diameter (D) and average distance (AD). The size of each node is proportional to its normalized degree. The outer nodes are parasites and the inner nodes are hosts

3.2. Model 2: fasciolosis spreading network (NW Spain)

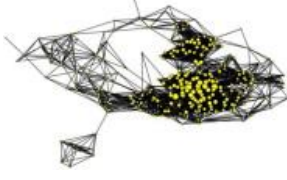
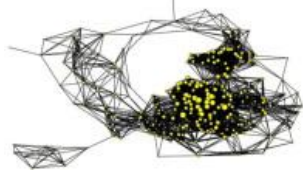
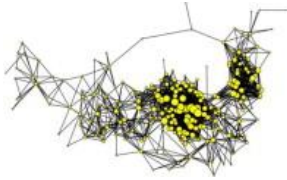
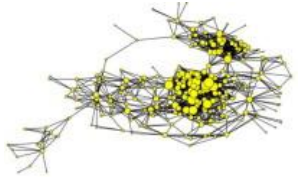
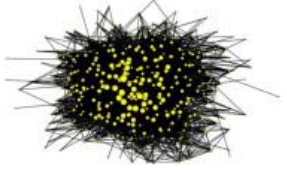
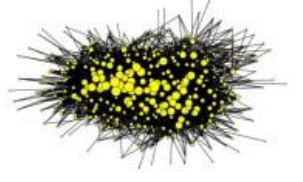
Fasciolosis is a parasitic infection caused by *Fasciola hepatica* (liver fluke) that has become an important cause of lost productivity in livestock worldwide. Considered a secondary zoonotic disease until the mid-1990s, human fasciolosis is at present emerging or re-emerging in many countries, including increase of prevalence as well as intensity and geographical expansion. In fact, research studies in recent years have justified the inclusion of fasciolosis in the list of important human parasitic diseases. At present, fasciolosis is the vector-borne disease showing the widest latitudinal, longitudinal and altitudinal distribution known. In addition, it shows a range of epidemiological characteristics related to a wide diversity of environments (Mas-Coma, 2005). In this sense, the study of geographical spreading of fasciolosis became a subject of great interest. In fact, in a recent work we have created a network (with two versions, **C** and **L**) to study the landscape spreading of fasciolosis in Galicia (NW Spain) (González-Díaz et al., 2010). However, we do not have quantitative criteria on the quality of the network connectivity, and re-sampling of all data in order to re-evaluate this connectivity in a field study is a hard and expensive task in terms of time and resources. This situation has prompted us to seek models in order to assess the quality of the networks previously assembled. The best QSPR models found for **C** and **L** were:

$$S(C_{ij}) = 2.86 \cdot [WC_1(f_i) - WC_1(f_j)] - 0.38, n = 24,163; \chi^2 = 4,148; p < 0.001 \quad (5)$$

$$S(L_{ij}) = 8.34 \cdot [WC_1(f_i) - WC_1(f_j)] - 2.17 \cdot [WC_5(f_i) - WC(f_j)] - 0.56, n = 23,991; \chi^2 = 1965; p < 0.001 \quad (6)$$

The values $WC_k(f_i)$ and $WC_k(f_j)$ used in this equation quantify information about the connectivity patterns between farms in the network. As it can be seen in the equations described in the Materials and, Methods Section, the connectivity of **C** depends on the spatial coordinates (x_i , y_i) of the farm (f_i), the altitude of the place (h_i), and the anti-parasite drug treatment (Tr_j) used to prevent fasciolosis on this farm. Consequently, the matrix **C** quantifies the *a priori* propensity $C_{ij} = 1$ of this disease to spread between farms immediately after treatment depending on geographical conditions. On the other hand, matrix **L** includes both criteria: (i) the preexistence of a high propensity of disease spreading $C_{ij} = 1$ and (ii) the experimental confirmation $L_{ij} = 1$ of a high risk ratio (RR_{ij}) of prevalence after treatment (PAT_j) for this disease on farms. The developed QSPR models show good values of accuracy, sensitivity, and specificity (see Table 1, Models 2a, 2b). Both observed and reconstructed networks (giant components) are described and represented graphically in Table 3.

Table 3. Comparison of observed vs. re-constructed NW Spain fasciolosis spreading networks and CoCoMac Brain Cortex network (giant components).

Observed network	Network descriptors ^a			Reconstructed Network
Fasciolosis spreading (C)				
	270	n	270	
	3048	m	3020	
	22.58	Ad	22.37	
	0.0839	den	0.0831	
	126.35	R	126.29	
	245,570	W	246,816	
	9	D	9	
	3.38	AD	3.40	
Fasciolosis spreading (L)				
	259	n	248	
	1798	m	1268	
	13.88	Ad	10.23	
	0.0538	den	0.0414	
	114.47	R	103.74	
	273,596	W	259,744	
	11	D	13	
	4.09	AD	4.24	
Macaque brain				
	360	n	357	
	5208	m	3829	
	28.93	Ad	21.45	
	0.0806	den	0.0603	
	145.98	R	118.60	
	292,892	W	295,056	
	5	D	5	
	2.27	AD	2.32	

^a Network descriptors: total number of connected nodes (n), number of edges (m), average degree (Ad), density (den), Randić index (R), Rucker index (W), diameter (D) and average distance (AD). The size of the drawn nodes is proportional to its normalized degree.

3.3. Model 3: cerebral cortex co-activation network

Connectivity is the key to understanding distributed and cooperative brain functions. Detailed and comprehensive data on large-scale connectivity between primate brain areas have been collated systematically from published reports of experimental tracing studies (Kotter, 2004). The eventual impact and success of connectivity databases, however, will require the resolution of several methodological problems that currently limit their use. These problems comprise four main points: (i) objective representation of coordinate-free, parcellation-based data; (ii) assessment of the reliability and precision of individual data, especially in the presence of contradictory reports; (iii) data mining and integration of large sets of partially redundant and contradictory data; and (iv) automatic and reproducible transformation of data between incongruent brain maps (Stephan et al., 2001). In order to address (ii) and (iv), we have developed a specific model for the collation of connectivity data on the macaque brain (CoCoMac) database (<http://www.cocomac.org>). The best model found was the following:

$$S(L_{ij}) = 1.92 \cdot WC_1(i) + 2.14 \cdot WC_2(j) - 1.68, n = 39,070; \chi^2 = 20,602; p < 0.001 \quad (7)$$

In this equation, $S(L_{ij})$ is a real-valued output variable that scores the propensity of the i th cerebral cortex region to undergo co-activation with the j th region in the CoCoMac network. The parameter χ_k quantifies the information related to the position of the i th region and their direct neighbors (j th regions) in the network after K steps. As in the previous equation the $\chi^2 = 4449$ statistics corresponds to a p -level < 0.001 , which indicates a significant discrimination between co-activated regions and not co-activated ones. The model shows very good values of accuracy, sensitivity, and specificity (see Table 1, Model 3). The giant components of the observed and reconstructed networks are described numerically and graphically in Table 3.

4. Conclusions

This work is aimed to introducing for the first time a new node descriptor, the Rucker–Markov invariants with two examples of network connectivity re-evaluation in Parasitology and Neuroinformatics: parasite–host networks, Fasciolosis spreading network (NW Spain) and Brain Cortex co-activation network. This new class of indices represents a modified version of the well-known Rucker's walk count (WC) index, by using Markov Chains, and they have been calculated with the MI-NODES tool. The descriptors have been used to create QSPR-like models in order to re-evaluate the links between the network nodes using the linear discriminant analysis. The overall accuracy for re-construction of complex networks of different Bio-Systems have been calculated: parasite–host networks (93.14%), NW Spain fasciolosis spreading networks (71.42/70.18%) and CoCoMac Brain Cortex co-activation network (86.40%). Therefore, we recommend the use of QSPR-like models based on Rucker-Markov descriptors as an alternative to high cost experimental re-evaluation of all links in any type of complex network.

Acknowledgments

Munteanu C. R. acknowledges Programme Isidro Parga Pondal, funded by Xunta de Galicia, Spain and European Social Fund (ESF). This work is supported by the Ibero-American Network of the Nano-Bio-Info-Cogno Convergent Technologies Network (Ibero-NBIC, 209RT-0366, CYTED, Spain) and by the “Neurogliale Artificial Network Processing of Information” project (TIN2009-07707, Ministerio de Ciencia e Innovación, Spain).

References

- Aguero-Chapin et al., 2008a. G. Aguero-Chapin, A. Antunes, F.M. Ubeira, K.C. Chou, H. Gonzalez-Diaz Comparative study of topological indices of macro/supramolecular RNA complex networks. *J. Chem. Inf. Model.*, 48 (2008), pp. 2265-2277.
- Aguero-Chapin et al., 2008b. G. Aguero-Chapin, H. Gonzalez-Diaz, G. de la Riva, E. Rodriguez, A. Sanchez-Rodriguez, G. Podda, R.I. Vazquez-Padron 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.*, 48 (2008), pp. 434-448.
- Aguero-Chapin et al., 2009. G. Aguero-Chapin, J. Varona-Santos, G.A. de la Riva, A. Antunes, T. Gonzalez-Villa, E. Uriarte, H. Gonzalez-Diaz Alignment-free prediction of polygalacturonases with pseudofolding topological indices: experimental isolation from *Coffea arabica* and prediction of a new sequence. *J. Proteome Res.*, 8 (2009), pp. 2122-2128.
- Aguiar-Pulido et al., 2010. V. Aguiar-Pulido, J.A. Seoane-Fernández, A.M. Freire-Veiga, H. Gonzalez-Diaz, A. Duardo-Sanchez, J. Dorado, A. Pazos, C.R. Munteanu New Markov-Randic Centralities for Computational methods of Biology, Parasitology, Technology, Social and Law networks. ICCMSE Kos, Greece (2010).
- Anderson and May, 1979. R.M. Anderson, R.M. May Population biology of infectious diseases. Part I. *Nature*, 280 (1979), pp. 361-367.
- Batagelj and Mrvar, 1998. V. Batagelj, A. Mrvar Pajek: a program for large network analysis. *Connections*, 21 (1998), pp. 47-57.
- Berca et al., 2011. M.N. Berca, A. Duardo-Sanchez, H. González-Díaz, A. Pazos, C.R. Munteanu Markov entropy for biology, parasitology, linguistic, technology social and law networks. H. González-Díaz, F.J. Prado-Prado, X. García-Mera (Eds.), *Complex Network Entropy: From Molecules to Biology, Parasitology, Technology, Social, Legal, and Neurosciences*, Transworld Research Network, Kerala, India (2011). 127-142.
- Boccaletti et al., 2006. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang Complex networks: structure and dynamics. *Physics Reports*, 424 (2006), pp. 175-308.
- Bonchev, 2003. D. Bonchev On the complexity of directed biological networks. *SAR QSAR Environ. Res.*, 14 (2003), pp. 199-214.
- Bonchev, 2004. D. Bonchev Complexity analysis of yeast proteome network. *Chem. Biodivers.*, 1 (2004), pp. 312-326.
- Bonchev and Buck, 2007. D. Bonchev, G.A. Buck From molecular to biological structure and back. *J. Chem. Inf. Model.*, 47 (2007), pp. 909-917.
- Bonchev and Rouvray, 2005. D. Bonchev, D.H. Rouvray Complexity in Chemistry, Biology and Ecology. Springer Science+Business Media, Inc., New York (2005).
- Bornholdt and Schuster, 2003. S. Bornholdt, H.G. Schuster Handbook of Graphs and Complex Networks: From the Genome to the Internet. WILEY-VCH GmbH & CO. KGa, Weinheim (2003).
- Casanola-Martin et al., 2006. G.M. Casanola-Martin, M.T. Khan, Y. Marrero-Ponce, A. Ather, M.N. Sultankhodzhaev, F. Torrens New tyrosinase inhibitors selected by atomic linear indices-based classification models. *Bioorg. Med. Chem. Lett.*, 16 (2006), pp. 324-330.
- Casanola-Martin et al., 2007. G.M. Casanola-Martin, Y. Marrero-Ponce, M.T. Khan, A. Ather, S. Sultan, F. Torrens, R. Rotondo TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: evaluation of different classification model combinations using bond-based linear indices. *Bioorg. Med. Chem.*, 15 (2007), pp. 1483-1503.
- Casanola-Martin et al., 2008. G.M. Casanola-Martin, Y. Marrero-Ponce, M. Tareq Hassan Khan, F. Torrens, F. Perez-Gimenez, A. Rescigno Atom- and bond-based 2D TOMOCOMD-CARDD approach and ligand-based virtual screening for the drug discovery of new tyrosinase inhibitors. *J. Biomol. Screen*, 13 (2008), pp. 1014-1024.
- Concu et al., 2009. R. Concu, M.A. Dea-Ayuela, L.G. Perez-Montoto, F. Bolas-Fernandez, F.J. Prado-Prado, G. Podda, E. Uriarte, F.M. Ubeira, H. Gonzalez-Diaz 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.*, 8 (2009), pp. 4372-4382.
- De Nooy et al., 2005. W. De Nooy, A. Mrvar, V. Batagelj Exploratory Social Network Analysis with Pajek. Cambridge University Press, Cambridge (2005).
- Dehmer and Emmert-Streib, 2009. M. Dehmer, F. Emmert-Streib. Analysis of Complex Networks: from Biology to Linguistics. Wiley-Blackwell, Weinheim (2009). 462 pp.

- Desdevises et al., 2002. Y. Desdevises, S. Morand, P. Legendre. Evolution and determinants of host specificity in the genus *Lamellogaster* (Monogenea). *Biol. J. Linn. Soc.*, 77 (2002), pp. 431-443.
- Detwiler and Janovy, 2008. J. Detwiler, J. Janovy Jr. The role of phylogeny and ecology in experimental host specificity: Insights from a eugregarine–host system. *J. Parasitol.*, 94 (2008), pp. 7-12.
- Duardo-Sanchez et al., 2011. A. Duardo-Sanchez, G. Patlewicz, H. González-Díaz. A review of network topological indices from chem.–bioinformatics to legal sciences and back. *Curr. Bioinf.*, 6 (2011), pp. 53-70.
- Estrada et al., 2010. E. Estrada, E. Molina, D. Nodarse, E. Uriarte. Structural contributions of substrates to their binding to P-glycoprotein. A TOPS-MODE approach. *Curr. Pharm. Des.*, 16 (2010), pp. 2676-2709.
- Estrada et al., 2006. E. Estrada, E. Uriarte, E. Molina, Y. Simon-Manso, G.W. Milne. An integrated in silico analysis of drug-binding to human serum albumin. *J. Chem. Inf. Model.*, 46 (2006), pp. 2709-2724.
- Estrada et al., 2001. E. Estrada, E. Uriarte, A. Montero, M. Teijeira, L. Santana, E. De Clercq. A novel approach for the virtual screening and rational design of anticancer compounds. *J. Med. Chem.*, 43 (2001), pp. 1975-1985.
- Estrada et al., 2002. E. Estrada, S. Vilar, E. Uriarte, Y. Gutierrez. In silico studies toward the discovery of new anti-HIV nucleoside compounds with the use of TOPS-MODE and 2D/3D connectivity indices. 1. Pyrimidyl derivatives. *J. Chem. Inf. Comput. Sci.*, 42 (2002), pp. 1194-1203.
- Gonzalez-Diaz, 2010a. H. Gonzalez-Diaz. Network topological indices, drug metabolism, and distribution. *Curr. Drug Metab.*, 11 (2010), pp. 283-284.
- Gonzalez-Diaz, 2010b. H. Gonzalez-Diaz. QSAR and complex networks in pharmaceutical design, microbiology, parasitology, toxicology, cancer, and neurosciences. *Curr. Pharm. Des.*, 16 (2010), pp. 2598-2600.
- Gonzalez-Diaz et al., 2010a. H. Gonzalez-Diaz, A. Duardo-Sanchez, F.M. Ubeira, F. Prado-Prado, L.G. Perez-Montoto, R. Concu, G. Podda, B. Shen. Review of MARCH-INSIDE and complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. *Curr. Drug Metab.*, 11 (2010), pp. 379-406.
- González-Díaz et al., 2003. H. González-Díaz, Y. Marrero, I. Hernandez, I. Bastida, E. Tenorio, O. Nasco, E. Uriarte, N. Castanedo, M.A. Cabrera, E. Aguila, O. Marrero, A. Morales, M. Perez. 3D-MEDNEs: an alternative “in silico” technique for chemical research in toxicology. 1. Prediction of chemically induced agranulocytosis. *Chem. Res. Toxicol.*, 16 (2003), pp. 1318-1327.
- González-Díaz et al., 2010. H. González-Díaz, M. Mezo, M. González-Warleta, L. Muñío-Pose, E. Paniagua, F.M. Ubeira. Network prediction of fasciolosis spreading in Galicia (NW Spain). H. González-Díaz, C.R. Munteanu (Eds.), *Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks*, Transworld Research Network, Kerala (India) (2010), pp. 191-204.
- González-Díaz and Munteanu, 2010. H. González-Díaz, C.R. Munteanu. *Topological indices for medicinal chemistry, biology, parasitology, neurological and social networks*. Transworld Research Network, Kerala, India (2010). pp. 1–212.
- Gonzalez-Diaz et al., 2011. H. Gonzalez-Diaz, F. Prado-Prado, X. Garcia-Mera, N. Alonso, P. Abeijon, O. Caamano, M. Yanez, C.R. Munteanu, A. Pazos, M.A. Dea-Ayuela, M.T. Gomez-Munoz, M.M. Garijo, J. Sansano, F.M. Ubeira. 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.*, 10 (2011), pp. 1698-1718.
- Gonzalez-Diaz et al., 2008. H. Gonzalez-Diaz, F. Prado-Prado, F.M. Ubeira. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr. Top Med. Chem.*, 8 (2008), pp. 1676-1690.
- Gonzalez-Diaz et al., 2010b. H. Gonzalez-Diaz, F.J. Prado-Prado, X. Garcia-Mera, N. Alonso, P. Abeijon, O. Caamano, M. Yanez, C.R. Munteanu, A. Pazos Sierra, M.A. Dea-Ayuela, M.T. Gomez-Munoz, M.M. Garijo, J. Sansano, F.M. Ubeira. 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 gallinae*. *J. Proteome Res.* (2010).
- Gonzalez-Diaz et al., 2010c. H. Gonzalez-Diaz, F. Romaris, A. Duardo-Sanchez, L.G. Perez-Montoto, F. Prado-Prado, G. Patlewicz, F.M. Ubeira. Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, applications, and legal issues. *Curr. Pharm. Des.*, 16 (2010), pp. 2737-2764.

- González-Díaz et al., 2007. H. González-Díaz, S. Vilar, L. Santana, E. Uriarte. Medicinal chemistry and bioinformatics – current trends in drugs discovery with networks topological indices. *Curr. Top. Med. Chem.*, 7 (2007), pp. 1025-1039.
- Gutman et al., 2001. I. Gutman, C. Rucker, G. Rucker. On walks in molecular graphs. *J. Chem. Inf. Comput. Sci.*, 41 (2001), pp. 739-745.
- Hatcher et al., 2006. J.M. Hatcher, J.T.A. Dick, A.M. Dunn. How parasites affect interactions between competitors and predators. *Ecol. Lett.*, 9 (2006), pp. 1253-1271.
- Hill and Lewicki, 2006. T. Hill, P. Lewicki. *STATISTICS Methods and Applications*. StatSoft, Tulsa (2006).
- Junker et al., 2006. B.H. Junker, D. Koschuetzki, F. Schreiber. Exploration of biological network centralities with CentiBiN. *BMC Bioinf.*, 7 (2006), p. 219.
- Kotter, 2004. R. Kotter. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics*, 2 (2004), pp. 127-144.
- Lukovits and Trinajstić, 2003. I. Lukovits, N. Trinajstić. Atomic walk counts of negative order. *J. Chem. Inf. Comput. Sci.*, 43 (2003), pp. 1110-1114.
- Marrero-Ponce et al., 2010. Y. Marrero-Ponce, G.M. Casanola-Martin, M.T. Khan, F. Torrens, A. Rescigno, C. Abad. Ligand-based computer-aided discovery of tyrosinase inhibitors. Applications of the TOMOCOMD-CARDD method to the elucidation of new compounds. *Curr. Pharm. Des.*, 16 (2010), pp. 2601-2624.
- Marrero-Ponce et al., 2007a. Y. Marrero-Ponce, M.T. Khan, G.M. Casanola-Martin, A. Ather, M.N. Sultankhodzhaev, R. Garcia-Domenech, F. Torrens, R. Rotondo. Bond-based 2D TOMOCOMD-CARDD approach for drug discovery: aiding decision-making in 'in silico' selection of new lead tyrosinase inhibitors. *J. Comput. Aided Mol. Des.*, 21 (2007), pp. 167-188.
- Marrero-Ponce et al., 2007b. Y. Marrero-Ponce, M.T. Khan, G.M. Casanola Martin, A. Ather, M.N. Sultankhodzhaev, F. Torrens, R. Rotondo. Prediction of tyrosinase inhibition activity using atom-based bilinear indices. *ChemMedChem*, 2 (2007), pp. 449-478.
- Mas-Coma, 2005. S. Mas-Coma. Epidemiology of fascioliasis in human endemic areas. *J. Helminthol.*, 79 (2005), pp. 207-216.
- Mezo et al., 2008. M. Mezo, M. Gonzalez-Warleta, J.A. Castro-Hermida, F.M. Ubeira. Evaluation of the flukicide treatment policy for dairy cattle in Galicia (NW Spain). *Vet. Parasitol.*, 157 (2008), pp. 235-243.
- Modha and Singh, 2010. D.S. Modha, R. Singh. Network architecture of the long-distance pathways in the macaque brain. *Proc. Natl. Acad. Sci. U.S.A.*, 107 (2010), pp. 13485-13490.
- Munteanu et al., 2009. C.R. Munteanu, J.M. Vazquez, J. Dorado, A.P. Sierra, A. Sanchez-Gonzalez, F.J. Prado-Prado, H. Gonzalez-Diaz. Complex network spectral moments for ATCUN motif DNA cleavage: first predictive study on proteins of human pathogen parasites. *J. Proteome Res.*, 8 (2009), pp. 5219-5228.
- Newman, 2003. M. Newman. The structure and function of complex networks. *SIAM Rev.*, 45 (2003), pp. 167-256.
- Newman et al., 2002. M.E. Newman, D.J. Watts, S.H. Strogatz. Random graph models of social networks. *Proc. Natl. Acad. Sci. U.S.A.*, 99 (2002), pp. 2566-2572.
- Poulin, 2010. R. Poulin. Network analysis shining light on parasite ecology and diversity. *Trends Parasitol.*, 26 (2010), pp. 492-498.
- Poulin et al., 2011. R. Poulin, B.R. Krasnov, D. Mouillot. Host specificity in phylogenetic and geographic space. *Trends Parasitol.*, 27 (2011), pp. 355-361.
- Price et al., 1986. P.W. Price, M. Westoby, B. Rice, P.R. Atsatt, R.S. Fritz, J.N. Thompson, K. Mobley. Parasite mediation in ecological interactions. *Annu. Rev. Ecol. Syst.*, 17 (1986), pp. 485-505.
- Puzyn et al., 2010. T. Puzyn, J. Leszczynski, M.T.D. Cronin. Recent advances in QSAR studies: methods and applications. J. Leszczynski (Ed.), *Challenges and Advances in Computational Chemistry and Physics*, Springer (2010). 423 pp.
- Ramos de Armas et al., 2004. R. Ramos de Armas, H. González-Díaz, R. Molina, E. Uriarte. Markovian Backbone Negentropies: molecular descriptors for protein research. I. Predicting protein stability in Arc repressor mutants. *Proteins*, 56 (2004), pp. 715-723.
- Ratti et al., 2010. C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, S.H. Strogatz. Redrawing the map of Great Britain from a network of human interactions. *PloS One*, 5 (2010), p. e14248.
- Riera-Fernandez et al., 2012. P. Riera-Fernandez, C.R. Munteanu, M. Escobar, F. Prado-Prado, R. Martin-Romalde, D. Pereira, K. Villalba, A. Duardo-Sanchez, H. Gonzalez-Diaz. 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.*, 293 (2012), pp. 174-188.

- Riera-Fernández et al., 2011. P. Riera-Fernández, C.R. Munteanu, N. Pedreira-Souto, R. Martín-Romalde, A. Duardo-Sanchez, H. González-Díaz. Definition of Markov–Harary invariants and review of classic topological indices and databases in biology, parasitology, technology, and social-legal networks. *Curr. Bioinf.*, 6 (2011), pp. 94-121.
- Rodríguez-Soca et al., 2010. Y. Rodríguez-Soca, C.R. Munteanu, J. Dorado, A. Pazos, F.J. Prado-Prado, H. Gonzalez-Diaz. Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein–protein interactions. *J. Proteome Res.*, 9 (2010), pp. 1182-1190.
- Rucker and Rucker, 2000. G. Rucker, C. Rucker. Walk counts, labyrinthicity, and complexity of acyclic and cyclic graphs and molecules. *J. Chem. Inf. Comput. Sci.*, 40 (2000), pp. 99-106.
- Rucker and Rucker, 2001. G. Rucker, C. Rucker. Substructure, subgraph, and walk counts as measures of the complexity of graphs and molecules. *J. Chem. Inf. Comput. Sci.*, 41 (2001), pp. 1457-1462.
- Santana et al., 2008. L. Santana, H. Gonzalez-Diaz, E. Quezada, E. Uriarte, M. Yanez, D. Vina, F. Orallo. Quantitative structure–activity relationship and complex network approach to monoamine oxidase a and B inhibitors. *J. Med. Chem.*, 51 (2008), pp. 6740-6751.
- Speck-Planche and Cordeiro, 2011. A. Speck-Planche, M.N.D.S. Cordeiro. Application of bioinformatics for the search of novel anti-viral therapies: rational design of anti-herpes agents. *Curr. Bioinf.*, 6 (2011), pp. 81-93.
- Speck-Planche et al., 2011a. A. Speck-Planche, L. Guilarte-Montero, R. Yera-Bueno, J.A. Rojas-Vargas, A. Garcia-Lopez, E. Uriarte, E. Molina-Perez. Rational design of new agrochemical fungicides using substructural descriptors. *Pest. Manag. Sci.*, 67 (2011), pp. 438-445.
- Speck-Planche et al., 2011b. A. Speck-Planche, V.V. Kleandrova, F. Luan, M.N. Cordeiro. Multi-target drug discovery in anti-cancer therapy: fragment-based approach toward the design of potent and versatile anti-prostate cancer agents. *Bioorg. Med. Chem.*, 19 (2011), pp. 6239-6244.
- Speck-Planche et al., 2011c. A. Speck-Planche, V.V. Kleandrova, J.A. Rojas-Vargas. QSAR model toward the rational design of new agrochemical fungicides with a defined resistance risk using substructural descriptors. *Mol. Divers.*, 15 (2011), pp. 901-909.
- Speck-Planche et al., 2009. A. Speck-Planche, M.T. Scotti, A. García-López, V.P. Emerenciano, E. Molina-Pérez, E. Uriarte. Design of novel antituberculosis compounds using graph-theoretical and substructural approaches. *Mol. Divers.*, 13 (2009), pp. 445-458.
- Stephan et al., 2001. K.E. Stephan, L. Kamper, A. Bozkurt, G.A. Burns, M.P. Young, R. Kotter. Advanced database methodology for the collation of connectivity data on the Macaque brain (CoCoMac). *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 356 (2001), pp. 1159-1186.
- Strogatz, 2001. S.H. Strogatz. Exploring complex networks. *Nature*, 410 (2001), pp. 268-276.
- Strogatz, 2005. S.H. Strogatz. Complex systems: Romanesque networks. *Nature (England)* (2005). 365-366.
- Thomas and Bonchev, 2010. S. Thomas, D. Bonchev. A survey of current software for network analysis in molecular biology. *Hum. Genomics*, 4 (2010), pp. 353-360.
- Vilar et al., 2005. S. Vilar, E. Estrada, E. Uriarte, L. Santana, Y. Gutierrez. 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.*, 45 (2005), pp. 502-514.
- Vina et al., 2009. D. Vina, E. Uriarte, F. Orallo, H. Gonzalez-Diaz. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. *Mol. Pharm.*, 6 (2009), pp. 825-835.
- Watts and Strogatz, 1998. D.J. Watts, S.H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393 (1998), pp. 440-442.