Current Bioinformatics. 2013;8(4): 429-437

S2SNet: a tool for transforming characters and numeric sequences into star network topological indices in chemoinformatics, bioinformatics, biomedical, and social-legal sciences

Cristian R. Munteanu[1], Alexandre L. Magalhães[2], Aliuska Duardo-Sánchez[1,3], Alejandro Pazos[1] and Humberto González-Díaz[4,5]

[1] *Department of Information and Communication Technologies, Computer Science Faculty, University of A Coruña, 15071 A Coruña, Spain*
[2] *REQUIMTE/University of Porto, Faculty of Science, Chemistry Department, 4169-007 Porto, Portugal*
[3] *Department of Special Public Law, Financial and Tributary Law Area, Faculty of Law, University of Santiago de Compostela (USC), 15782, Spain*
[4] *Department of Organic Chemistry II, University of the Basque Country, UPV/EHU, 48940, Leioa, Spain*
[5] *IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

Abstract
The study of complex systems such as proteins/DNA/RNA or dynamics of tax law systems can be carried out with the complex network theory. This allows the numerical quantification of the significant information contained by the sequences of amino acids, nucleotides or types of tax laws. In this paper we describe S2SNet, a new Python tool with a graphical user interface that can transform any sequence of characters or numbers into series of invariant star network topological indices. The application is based on Python reusable processing procedures that perform different functions such as reading sequence data, transforming numerical series into character sequences, changing letter codification of strings and drawing the star networks of each sequence using Graphviz package as graphical back-end. S2SNet was previously used to obtain classification models for natural/random proteins, breast/colon/prostate cancer-related proteins, DNA sequences of mycobacterial promoters and for early detection of diseases and drug-induced toxicities using the blood serum proteome mass spectrum. In order to show the extended practical potential of S2SNet, this work presents several examples of application for proteins, DNA/RNA, blood proteome mass spectra and time evolution of the financial law recurrence. The obtained topological indices can be used to characterize systems by creating classification models, clustering or pattern search with statistical, Neural Network or Machine Learning methods. The free availability of S2SNet, the flexibility of analyzing diverse systems and the Python portability make it an ideal tool in fields such as Bioinformatics, Proteomics, Genomics, and Biomedicine or Social, Economic and Political Sciences.

Keywords:
Complex network; Financial law network; Graph indices; Interaction; Network; Protein; Python application; Social network

1. INTRODUCTION

The complexity of the real systems makes the comparison difficult between each other or the extraction of specific information that describes a discrete property. A possible strategy is to use the information about the connections or relationships between different parts of the entire system. This can be carried out using the Graph or Complex Network (CN) theory. A network is a collection of nodes represented by graphs composed of any items that have a) chemical, b) biological, c) social, and/or d) technological nature. For instance, atoms, molecules (proteins, DNA/RNA) are nodes of type a); viruses, bacteria, organisms, social actors and/or laws to regulate the behaviour of such actors may be considered as type b) and/or c); whereas computers, electric power plants, airports, mass spectra signals, or links between web pages or computers are usually allocated within the type d) of nodes [1-6]. This work presents four types of CN but we do not exclude other types of nodes and networks. In this approach we may use different invariant numbers often called Topological Indices (TIs) in order to encode or describe the structure of the systems. These TIs are usually derived from node-node adjacency or other types of matrices associated to the CN [7, 8]. Even though many TIs have been described only for molecular graphs of type a) many of them have been extended to be used in all types of CN [5, 9-13]. In any case, the development of new types of CN and graph representations or new TIs to describe them is an emerging field of science.

In the present work, a new tool for calculating TIs of a special type of CN is presented. S2SNet – Sequence to Star Network [14] is a free Python desktop application for Microsoft Windows XP/Vista operating system. This new software has a friendly graphical user interface made with wXPython [15]. The software can be used to transform sequences of characters into a CN with Star Network (SN) topology. The SNs were introduced by Randic *et al.* in order to analyze the protein sequences, but have been used to investigate DNA/RNA or MS spectra of the blood proteome [16, 17] and may be extended to several other sequence-type data such as music, text, time series, etc. In this approach, the network/graph matrices are translated into the DOT language and plotted with Graphviz (*twopi*, *neato*, *dot*, *circo* and *fdp*) [18]. This paper introduces S2SNet as free software and gives detailed examples on how to use it in order to transform a text or numeric sequence into different types of SNs, visualize the resulted SNs, and calculate different classes of TIs for these SNs. The examples of sequences given here are the following: 1) the primary structure of the proteins, DNA or RNA and 2) the intensity of the Mass Spectra signals and 3) the time evolution of the recurrence to different types of laws in a financial law system. Thus, S2SNet can be used to describe systems in several fields such as social, economical, or political sciences, bioinformatics, structural biology, and clinical proteomics.

A previous application, March-Inside [19], has been used to calculate topological indices for Spiral and Star Graphs. However, that tool can generate Spiral Graphs from a list of sequences and Star Graphs only from the graphical interface, one by one. Other softwares such as Centibin [20] and Pajek [21] are calculating some of the S2SNet topological indices and process file by file in a matrix format (mat/net). Thus, S2SNet has the following advantages: transforms a list of sequences into Star Graph TIs at the click of a mouse, has the possibility to generate embedded graphs, calculates an extended list of Star Graph topological indices, and plots the resulted graphs in several Graphviz formats.

2. SOFTWARE DESCRIPTION

S2SNet is a free Python application that can transform character sequences into topological Star Network indices, transform number series into sequences, transform an N-character sequence into a 1-character sequence by changing the codification, edit/view the input/output txt files, create DOT language files, plot and display the networks as PNG images. Its basic architecture is described in Fig. (**1**).

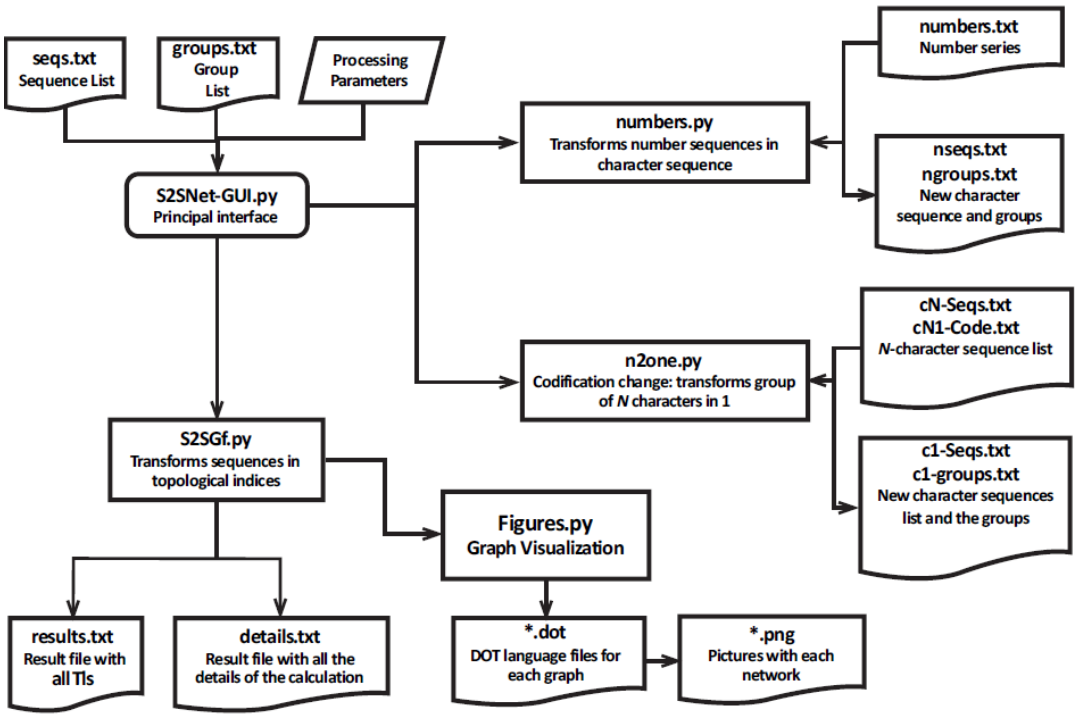


Fig. (1). Architecture of the S2SNet software: the tool has Python processing functions in, wxPython user interface and network representations created with the Graphviz package as graphical back-end.

S2SNet has two main panels, the principal window and the console output (see Fig. **2**). The main window has buttons for a fast access to the main features of the application:

– *S2SNet* - the transformation of sequences in SN TIs;

– *Help* - a short help page;

– *About* - details about S2SNet and the authors;

– *Quit* - leave the tool.

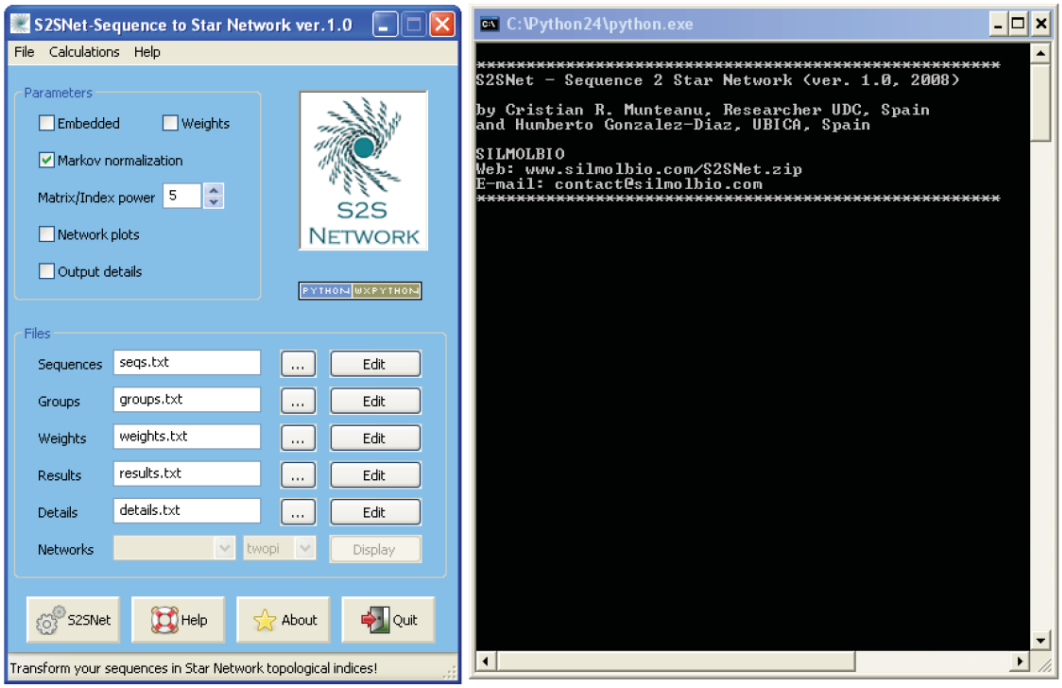


Fig. (2). S2SNet user interface.

The same options available in the S2SNet Menu divided into File, Calculations and Help.

In addition, the File and Calculations menus contain extra options:

– *New* – creates a new text document using the native Notepad from MS Windows;

– *Numbers to Sequence* – transforms the numerical input data such as the protein mass spectra into a 1-character sequence;

– *N to 1-Character Sequence* – changes the sequence codification by transforming N-character groups into 1- character sequences, such as the DNA/RNA codon sequences.

In the console panel, the details about the stage/errors/results of the calculations are displayed. In the main window you can choose the TIs calculation parameters, the input/output files and the visualization type of the resulted graphs:

Parameters: embedded network, the use of the weight for each character; Markov normalization of the connectivity matrixes; if you want details of the calculations containing all the intermediate matrices and other info; power of the connectivity matrices and of some indices (max. 5); networks plotting support;

– Input files: sequences, groups and weights files;

– Output files: results and details files;

– Display mode for the Network plots: the sequence to display and the type of drawing application (*dot*, *circo*, *twopi*, *neato* and *fdp*); extra theoretical graphs are calculated and plotted such as the maximum and the average graphs of the sequences introduced as input.

All the options have default values for a calculation characterized by a non-embedded graph, a Markov normalizations of the matrices, a power of 5 for the matrices, no weight for the nodes, no network plot support and no detail file. In the displayed graphs, each group has a different color. If you need to obtain modified plots, you can find the DOT files (one for each sequence) and the Graphviz executables (*dot*, *circo*, *twopi*, *neato*, *fdp*) in the "dot" folder (if you enable the Network plot option).

The Calculation menu allows you to transform your data into the S2SNet format (1-character string):

– *Numbers to Sequence* (see Fig. **3**) - transforms your numbers into a sequence (the numbers must be TAB separated); you can choose the following parameters:

- The minimum and the maximum values of your data, number of groups you need (a maximum of 80); you have a GET button if you want to use the minimum and maximum calculated from your entire data;
- The input files: number (data) file;
- The output files: sequence files, group file and interval file (description of the group range).

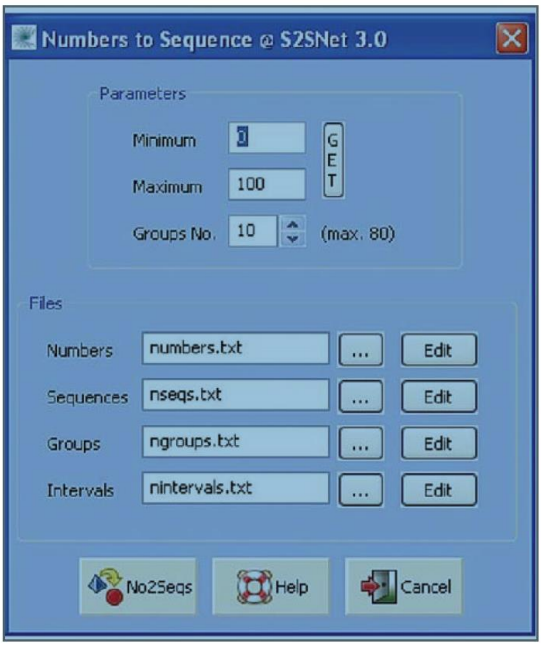


Fig. (3). Numbers to Sequence filter in S2SNet.

This filter can be used to transform the protein mass spectra numbers into text sequences and to calculate the corresponding Star Network indices.

– *N to 1-Character Sequence* (see Fig. **4**) - Transform your N-character sequences into 1-character sequences; you can set the following:

- The input files: N-character file (initial file), code file (the equivalence between N-character and 1- character; ex: ALA=A);
- The output files: 1-character files for S2SNet (final file) and group file (one item groups).

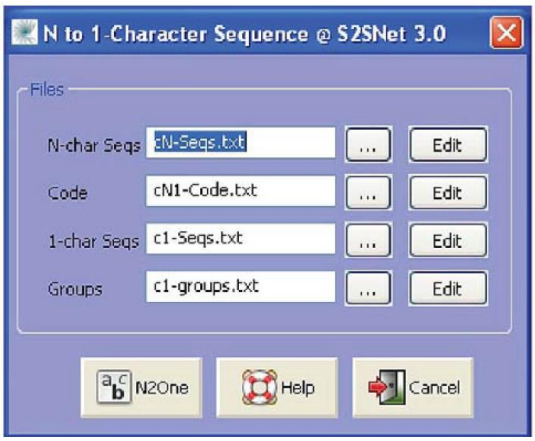


Fig. (4). N to 1-Character Sequence filter in S2SNet.

This filter can be used to transform sequences that contain items described by 3-letter codes such as the amino acids 3-letter code into sequences of 1-letter code; another example is the translation of the 3-nucleotides codons into the corresponding 1-letter amino acid sequence.

3. TIs OF SNs

In order to explain what an SN is, the protein representation is used in this section. Thus, a complex network can have the amino acids as vertices (nodes), connected in a specific sequence by the peptide bonds. The theoretical abstraction of a network is represented by a graph. The star graph is a special case of trees with N vertices where one has got N-1 degrees of freedom and the remaining N-1 vertices have got one single degree of freedom [22]. In the case of proteins, each of the 20 possible branches ("rays") of the star contains the same amino acid type and the star center is a non-amino acid vertex. A protein primary sequence can be represented by different forms of graphs, which can be associated with distinct distance matrices (Randic *et al.*, 2007). The best method to construct a standard star graph is the following: each amino acid/vertex holds the position in the original sequence and the branches are labeled in the alphabetical order of the 3-letter amino acid code [17]. The embedded graph contains the initial sequence connectivity in the protein chain. Fig. (**5**) presents the non-embedded (A) and the embedded (B) star graphs of the HIV gp120 C5 protein (1meq: VKIEPLGVAPTKAKRRVVQREKR) using the alphabetical order of one-letter amino acid code. Thus, the primary structure of protein chains is transformed into the corresponding star graph invariant TIs. The resulted graphs do not depend on the 3-dimensional structure or the shape of the protein. The derived connectivity matrix, distance matrix and degree matrix are used to compare the graphs/networks. The matrices of the connectivity in the sequence and in the star graph are combined in the case of the embedded graph. These matrices and the normalized ones form the base for the TIs calculation.

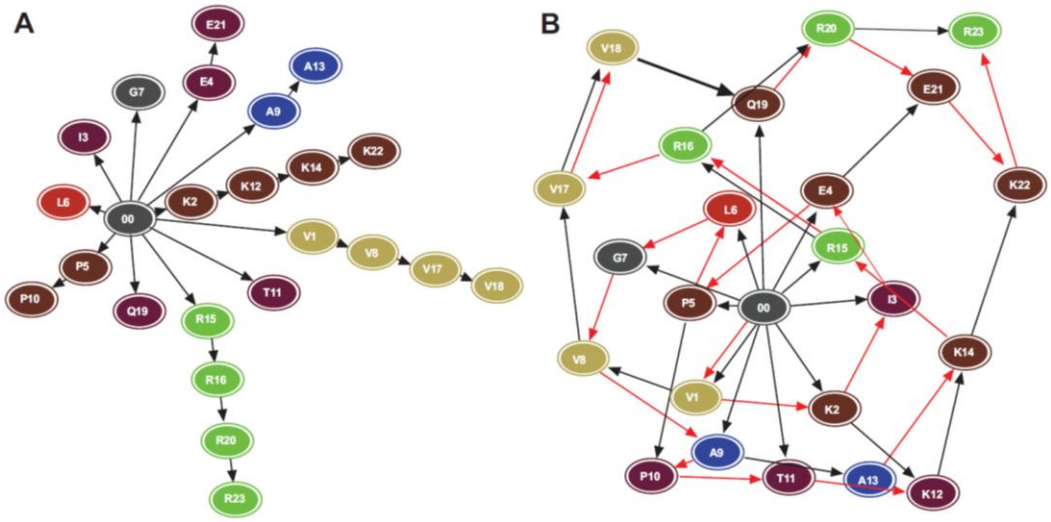


Fig. (5). The non-embedded (**A**) and embedded (**B**) graphs of HIV gp120 C5 protein.

The algorithm is begun by reading the sequences, groups and weights (see Fig. **1**). The transformation of each sequence into an SN is translated into a connectivity matrix (*M*), node degree vector (*deg*) and distance node matrix (*d*). If the graph is non-embedded *M* it includes only the modified connectivity inside the SG. In the case of the embedded graphs, the original sequence connectivity will be added. The node degree represents the number of connections for a node in the graph and the distance matrices is filled with the number of nodes between each pair of nodes along the graph connections. If the weights option is chosen, *M* will have the corresponding weight values along the matrix diagonal. In the Markov normalization (default option), *M* is first normalized by dividing each element to the sum of the elements by row and after that raising it to the power *n* given by the user (default is 5), resulting *n* matrices (M^n). In case of non-Markov normalization, in the first step the matrix is powered, resulting *n* matrices that are normalized by the same division.

These variables form the base for the calculation of the following TIs, presented in the output file (Todeschini and Consonni, 2002):

– Shannon Entropy of the *n* Markov Matrices (*Sh*):

$$Sh_n = -\sum_i p_i * \log(p_i) \quad (1)$$

where p_i are the n_i elements of the vector *p* resulting from the matrix multiplication of the powered Markov normalized matrix (n_i x n_i) and a vector (n_i x 1) with each element equal to $1/n_i$;

– Trace of the *n* connectivity matrices (*Trn*) or the spectral moments:

$$Tr_n = \sum_i (M^n)_{ii} \quad (2)$$

where n = 0 – power limit, M = graph connectivity matrix (i*i dimension); ii = i^{th} diagonal element;

– -Harary number (*H*) or the reciprocal distance sum index:

$$H = \sum_{i<j} m_{ij}/d_{ij} \quad (3)$$

where d_{ij} are the elements of the distance matrix and m_{ij} are the elements of the M connectivity matrix;

– Wiener index (*W*) or the sum of the numbers of edges in the shortest paths in a graph between all pairs of amino acids in a protein:

$$W = \sum_{i<j} d_{ij} \tag{4}$$

– Gutman topological index (S_6):

$$S_6 = \sum_{ij} deg_i * deg_j / d_{ij} \tag{5}$$

where deg_i are the elements of the degree matrix;

– Schultz topological index (non-trivial part) (*S*):

$$S = \sum_{i<j} \left(deg_i + deg_j\right) * d_{ij} \tag{6}$$

– Balaban distance connectivity index (*J*) or average distance sum connectivity index (measures the graph ramification):

$$J = edges/(edges - nodes + 2) * \sum_{i<j} m_{ij} * sqrt\left(\sum_k d_{ik} * \sum_k d_{kj}\right) \tag{7}$$

where nodes+1 = AA numbers/node number in the Star Graph + origin, $\sum_k d_{ik}$ is the node distance degree;

– Kier-Hall connectivity indices (^{n}X):

$$^0\text{X} = \sum_i 1/sqrt(deg_i) \tag{8}$$

$$^2\text{X} = \sum_{i<j<k} m_{ij} * m_{jk} / sqrt(deg_i * deg_j * deg_k) \tag{9}$$

$$^3\text{X} = \sum_{i<j<k<m} m_{ij} * m_{jk} * m_{km} / sqrt(deg_i * deg_j * deg_k * deg_m \tag{10}$$

$$^4\text{X} = \sum_{i<j<k<m<o} m_{ij} * m_{jk} * m_{km} * m_{mo} / sqrt(deg_i * deg_j * deg_k * deg_m * deg_o) \tag{11}$$

$$^5\text{X} = \sum_{i<j<k<m<o<q} m_{ij} * m_{jk} * m_{km} * m_{mo} * m_{oq} / sqrt(deg_i * deg_j * deg_k * deg_m * deg_o * deg_q) \tag{12}$$

– Randic connectivity index (^{1}X):

$$^1\text{X(R)} = \sum_{ij} m_{ij} / sqrt(deg_i * deg_j) \tag{13}$$

These indices can be used in the next step in order to construct any classification model, clustering or pattern search using the statistical, neural networks or machine learning methods from applications such as STATISTICA [23] or WEKA [24].

4. RESULTS AND DISCUSSION

S2SNet has been successfully used by our group in four previous papers in order to create classification models for evaluating a protein as natural or random [25], as breast/colon [26] or prostate [27] cancer-related, a DNA sequence as a Mycobacterial promoter [28], and for early detection of diseases and drug-induced toxicities using the blood serum proteome mass spectrum [29]. This work presents four cases of S2SNet calculations for proteins, DNA, mass spectra and laws.

In the first case, the primary structure information of a protein sequence (the amino acid order and type) is encrypted in SN TIs. Fig. (**6**) shows the resulted graphs for 7ODC, chain A from the Protein Data Bank [30] obtained with four Graphwiz tools from S2SNet. The nodes represent the protein amino acids linked by the peptide bonds and are grouped into 20 branches corresponding to the different natural amino acids. The resulted TIs of the embedded SN are presented in Table **1**.

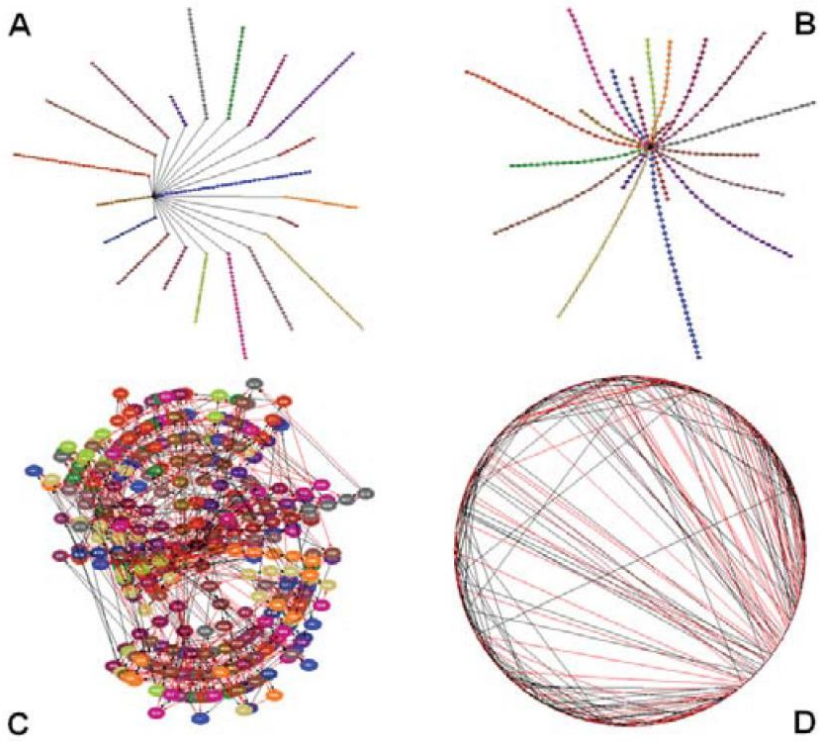
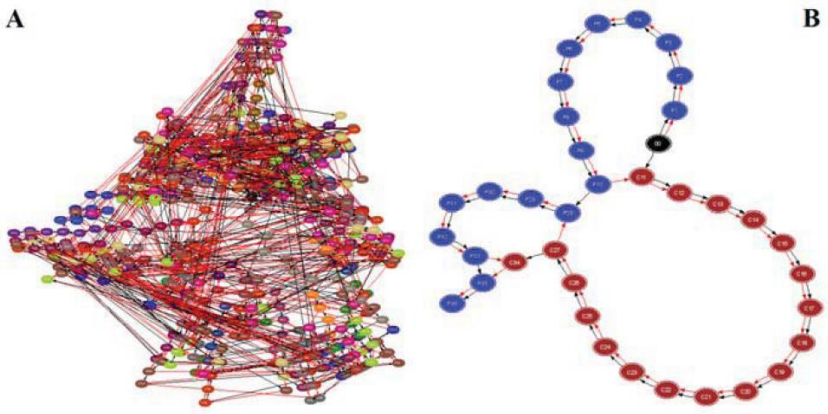


Fig. (6). 7ODCA star graphs: non-embedded SN created with twopi (**A**) and neato (**B**) and embedded SN created with twopi (**C**) and circo (**D**).

Table 1. The TIs of the Embedded SNs for the Protein 7ODC, the Gene COMT, and the Mass Spectrum of a Blood Proteome Sample of a Cancer Patient

TI	7ODC Chain A Protein	COMT Gene	Blood Proteome Mass Spectrum
Sh0	5.94	6.63	3.59
Sh1	6.20	6.82	3.68
Sh2	6.15	6.79	3.67
Sh3	6.21	6.84	3.69
Sh4	6.19	6.83	3.68
Sh5	6.21	6.85	3.69
Tr0	388.00	769.00	37.00
Tr1	0.00	0.00	0.00
Tr2	101.03	202.65	17.22
Tr3	1.48	6.19	0.22
Tr4	46.84	95.93	12.27
Tr5	2.30	7.77	0.35
H	2149.33	4785.45	118.46
W	9735114.00	75792640.00	8436.00
S6	64058.51	140090.82	1113.23
S	75274220.00	578800473.00	36536.00
J	77495217.40	604518229.56	145898.28
X0	199.40	396.63	25.55
X1(R)	192.71	382.74	18.36
X2	184.62	367.67	13.04
X3	176.73	353.05	9.12
X4	168.77	338.93	6.27
X5	160.91	325.34	4.31

The next application is the DNA codons of COMT gene that is virtually translated with S2SNet into an amino acid sequence. This sequence is represented as an SN with the amino acids as nodes (equivalent to the codons) distributed in 21 branches, 20 standard amino acids and an extra X nonamino acid corresponding to the STOP DNA codons [31]. Thus, the primary structure of a DNA segment is transformed into SN TIs. The corresponding star network is presented in Fig. (**7A**). COMT (catechol-O-methyltransferase) is a gene that controls the function of the catechol-Omethyltransferase enzyme. This enzyme metabolizes catecholamines, which are heavily linked to dopaminergic and adrenergic/noradrenergic neurotransmission, or endorphins [32]. The variations of the COMT gene known as Single Nucleotide Polymorphism (SNP or SNIP) is thought by some researchers to be one of the key genes associated with schizophrenia [33].



Fig, (7). Embedded SNs for the COMT gene (**A**) created with twopi and for the blood serum mass spectrum (**B**) carried out with neato.

Another example is the blood serum proteome mass spectrum that contains the patient's information about positive drug induced toxicity. They are then transformed into a sequence of characters (nodes), which define the group of SN upon the different range values of the signal intensity. The initial connectivity is generated by the signal positions in the spectrum. The embedded SNs of the COMT gene and proteome mass spectra are included in Fig. (**7**), and the corresponding embedded TIs are in Table **1**.

The last example of the S2SNet use is dedicated to the Social Network Analysis (SNA). SNA may be defined as the disciplined inquiry into patterning of relations among social actors, as well as the patterning of relationships among actors at different levels of analysis (such as persons and groups) [34]. It provides a common approach for all those disciplines involved in social structure study [35-38] susceptible of network depiction. Social structure concept is merely used in sociology and social theory. Although there is no agreement between theorists, it can refer to a specific type of relation between entities or groups, it can also evolve enduring patterns of behaviour and relationship within a society, or social institutions and regulations becoming embedded into social systems. For the most comprehensive review of SNA see the in-depth review of Newman M entitled *The Structure and Function of Complex Networks* [39]. In any case, considering that a network is a set of items, usually called *nodes*, with connections between them, socalled *edges*, then we have a representation of social relationships in terms of nodes and ties, where nodes can be the individual actors within the networks, and ties the relationships between these actors [3]. In fact, SNA is nothing new in social sciences studies, as in the early 1930's, sociologists had already created a social network to study friendships between school children [40]. Since then, the importance of network approach to social sciences increased dramatically, and its applications expended from interrelation between family members [41] to company business interaction [42, 43] or patterns of sexual contacts [44, 45]. Although the network approach is so pervasive in the social sciences, its application in the law scope is still weak. Network tools and methodologies might be useful to illustrate the interrelation between different law types, and check the importance of a specific instrument so as the normative hierarchy respected by legislators. This helps to regulate the most important matter for individuals through law instruments which requires the approval from the most representative democratic actors. In this sense, the S2Snet software is a novel tool which has enabled herein the representation of the basic laws of the Spanish tax law system. Considering all this, we have built a graph on the recurrence to different types of laws related to tax matters over the years since 1946, see Fig. (**8**).

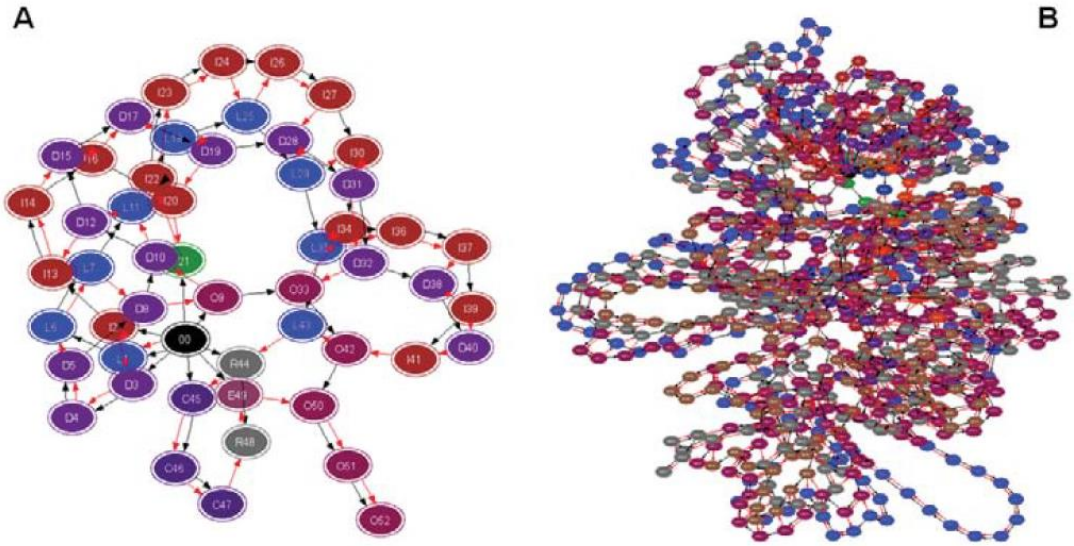


Fig. (8). Embedded SN for Spanish Financial law system over two periods: **A**) from 1946 to 1978 and **B**) from 1946 to 2004 (cumulative behavior with respect to 1946) using the neato algorithm.

In these financial laws SN, we represented the time series for the use of different laws as a one-letter code sequence. In this sequence each specific type of rule is represented by a one letter code. Specifically, we transform all the regulation types as follows: Law (L), Instrument of Ratification (I), Decree-law (D), Order (O), Protocol (P), Royal Decree (R), Circular (C), EC Council Directive (E), Resolution (S), Organic Law (G), Agreement (A), Instruction (T), Convention (V), and Regulation (M). From there, a stargraph which connects several branches to a central node may be built. Each branch is composed of the same type of laws, and nodes -whether on the same branch or another being connected to each other if two laws are used at a time one after another. This method describes numerically the recurrence to different regulations or group of regulations or laws and can be used to describe the past dynamics and predict the future behaviour of the tax law application in Spain or other countries. In any case, many potential implications are still to be discovered in future research beyond this introductory work. In Table **2**, we illustrate the behaviour of different TIs for the SN constructed that reflect the changes in the recurrence to different laws in these periods.

Table 2. TIs of the Embedded SN for Spanish Financial Law System Over Two Periods: A) from 1946 to 1978 and B) from 1946 to 2004 (Cumulative Behavior with Respect to 1946), see also Fig. (8)

TI	1946-1978	1946-2004
Sh0	3.93	6.77
Sh1	4.13	6.96
Sh2	4.12	6.97
Sh3	4.17	7.01
Sh4	4.17	7.02
Sh5	4.19	7.04
Tr0	53.00	898.00
Tr1	0.00	0.00
Tr2	16.10	274.25
Tr3	1.56	12.50
Tr4	8.89	149.29
Tr5	1.90	16.48
H	188.52	5727.39
W	24804.00	120691649.00
S6	4380.84	130175.86
S	167069.00	812264716.00
J	192361.90	997661638.98
X0	29.67	499.56
X1(R)	26.14	445.92
X2	22.05	397.97
X3	18.65	355.59
X4	15.93	318.33
X5	13.65	284.83

5. CONCLUSIONS

This paper is proposing the use of S2SNet as a new tool in the studies of the SN calculating the TIs for characters or numbers sequences. It is a fast and free application that generates the invariant TI set for a specific sequence. These TIs can be used as the base for the development of Statistical, Neural Networks, or Machine Learning classification models/clustering/pattern search models. Thus, this tool can be used for e-learning or research in different fields such as Bioinformatics, Proteomics, Genomics, Biomedicine or Social, Economic, and Political Sciences. In this paper, we have presented only a few examples of possible SNs, but the application of S2SNet is not limited to any type of sequence.

CONFLICT OF INTEREST

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

ACKNOWLEDGEMENTS

Cristian R. Munteanu acknowledges the funding support of the “Isidro Parga Pondal” Program from Xunta de Galicia (supported by the European Social Fund).

REFERENCES

[1] Barabasi AL. Sociology. Network theory--the emergence of the creative enterprise. Science 2005; 308(5722): 639-41.
[2] Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004; 5(2): 101-13.
[3] Bornholdt S, Schuster HG. Handbook of Graphs and Complex Networks: From the Genome to the Internet. WILEY-VCH GmbH & CO. KGa: Wheinheim 2003.
[4] González-Díaz H, Vilar S, Santana L, Uriarte E. Medicinal chemistry and bioinformatics – current trends in drugs discovery with networks topological indices. Curr Top Med Chem 2007; 7(10): 1025-39.
[5] Mason O, Verwoerd M. Graph theory and networks in Biology. IET Syst Biol 2007; 1(2): 89-119.
[6] Yook SH, Jeong H, Barabasi AL. Modeling the Internet's largescale topology. Proc Natl Acad Sci USA 2002; 99(21): 13382-6.
[7] García-Domenech R, Gálvez J, Julián-Ortiz JV, Pogliani L. Some new trends in chemical graph theory. Chem Rev 2008; 108(3): 1127–69.
[8] Gonzalez-Diaz H, Gonzalez-Diaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. Proteomics 2008; 8(4): 750-78.
[9] Concu R, Podda G, Uriarte E, González-Díaz H. A new computational chemistry & complex networks approach to structure - function and similarity relationships in protein enzymes. In: Handbook of Computational Chemistry Research; Collett CTaR, CD., ed., Nova Science Publishers 2009.
[10] Dall'asta L, Alvarez-Hamelin I, Barrat A, Vazquez A, Vespignani A. Statistical theory of Internet exploration. Phys Rev E Stat Nonlin Soft Matter Phys 2005; 71(3 Pt 2A): 036135.
[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] Gonzalez-Diaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. Curr Top Med Chem 2008; 8(18): 1676-90.
[13] Todeschini R, Consonni V. Handbook of Molecular Descriptors. Wiley-VCH 2002.
[14] Munteanu CR, Gonzáles-Diáz H. S2SNet - Sequence to Star Network, Reg. No. 03 / 2008 / 1338, Santiago de Compostela, Spain. Santiago de Compostela, Spain 2008.
[15] Rappin N, Dunn R. wxPython in Action. Manning Publications Co.: Greenwich, CT 2006.
[16] Ferino G, Delogu G, Podda G, Uriarte E, González-Díaz H. Quantitative Proteome-Disease Relationships (QPDRs) in Clinical Chemistry: Prediction of Prostate Cancer with Spectral Moments of PSA/MS Star Networks. In: Clinical Chemistry Research; Mitchem BHaS, Ch.L., ed., Nova Science Publisher: NY 2009.
[17] Randic M, Zupan J, Vikic-Topic D. On representation of proteins by star-like graphs. J Mol Graph Model 2007: 290-305.
[18] Koutsofios E, North SC. Drawing Graphs with dot. AT&T Bell Laboratories, Murray Hill: NJ, USA 1993.
[19] González-Díaz H, Molina-Ruiz R, Hernandez I. March-Inside version 3.0 (Markov Chains Invariants for Simulation & Design); Windows supported version under request to the main author contact email: gonzalezdiazh@yahoo.es. 3.0 ed 2007.
[20] Koschützki D. CentiBiN Version 1.4.2. 2006: CentiBiN Version 1.4.2, Centralities in Biological Networks © 2004-6 Dirk Koschützki Research Group Network Analysis, IPK Gatersleben, Germany.
[21] Batagelj V, Pajek M.A. Program for Large Network Analysis (ver. 1.15), http://vlado.fmf.uni-lj.si/pub/networks/pajek/. 1.15 ed 2006.
[22] Harary F. Graph Theory. Westview Press: MA 1969.

[23] StatSoft.Inc. STATISTICA, (data analysis software system), version 6.0, www.statsoft.com. 6.0 ed 2002.
[24] Witten IH, Frank E. WEKA: Waikato Environment for Knowledge Analysis. 2000.
[25] [25] Munteanu CR, Gonzalez-Diaz H, Borges F, de Magalhaes AL. Natural/random protein classification models based on star network topological indices. J Theor Biol 2008; 254(4): 775-83.
[26] Munteanu CR, Magalhaes AL, Uriarte E, Gonzalez-Diaz H. Multitarget QPDR classification model for human breast and colon cancer-related proteins using star graph topological indices. J Theor Biol 2009; 257(2): 303-11.
[27] González-Díaz H, Ferino G, Munteanu CR, Vilar S, Uriarte E. Protein Graphs in Cancer Prediction. In: Oncoproteomics; Cho WK, ed., Springer 2009.
[28] Perez-Bello A, Munteanu CR, Ubeira FM, De Magalhaes AL, Uriarte E, Gonzalez-Diaz H. Alignment-free prediction of mycobacterial DNA promoters based on pseudo-folding lattice network or star-graph topological indices. J Theor Biol 2009; 256(3): 458-66.
[29] Cruz-Monteagudo M, Munteanu CR, Borges F, *et al.* Stochastic molecular descriptors for polymers. 4. Study of complex mixtures with topological indices of mass spectra spiral and star networks: The blood proteome case. Polymer 2008; 49: 5575-87.
[30] Berman HM, Westbrook J, Feng Z, *et al.* The Protein Data Bank. Nucleic Acids Res 2000; 28: 235-42.
[31] Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. Introduction to Genetic Analysis. 7th ed. W. H. Freeman & Co.: New York 1999.
[32] Zubieta JK, Heitzeg MM, Smith YR, *et al.* COMT val158met genotype affects mu-opioid neurotransmitter responses to a pain stressor. Science 2003; 299(5610): 1240-3.
[33] Wonodi I, Mitchell BD, Stine OC, *et al*. Lack of association between COMT gene and deficit/nondeficit schizophrenia. Behav Brain Funct 2006; 2: 42.
[34] Breiger R. The Analysis of Social Networks. In: Handbook of Data Analysis; Hardy M, Bryman A, eds., Sage Publications: London 2004; 505-26.
[35] Abercrombie N, Hill S, Turner BS. The Penguin Dictionary of Sociology. In: Social structure, Penguin: London 2000.
[36] Craig C. Social Structure, Dictionary of the Social Sciences, Oxford University Press: Oxford 2002.
[37] Wellman B, Berkowitz SD. Social Structures: A Network Approach. Cambridge University Press: Cambridge 1988.
[38] White H, Boorman S, Breiger R. Social Structure from Multiple Networks: Blockmodels of Roles and Positions. American Journal of Sociology 1976; 81(730-780).
[39] Newman M. The structure and function of complex networks. SIAM Review 2003; 56: 167-256.
[40] Moreno JL. Who Shall Survive?. Beacon House: New York 1934.
[41] Padgett JF, Ansell CK. Robust Action and the Rise of the Medici, 1400-1434. Amer J Sociol 1993; 98(6): 1259-319.
[42] Mariolis P. Interlocking directorates and control of corporations: The theory of bank control. Social Sci Quart 1975; 56: 425-39.
[43] Mizruchi MS. The American Corporate Network, 1904-1974. Sage: Beverly Hills 1982.
[44] Klovdahl AS, Potterat JJ, Woodhouse DE, Muth JB, Muth SQ, Darrow WW. Social networks and infectious disease: the Colorado Springs Study. Soc Sci Med 1994; 38(1): 79-88.
[45] Liljeros F, Edling CR, Amaral LA, Stanley HE, Aberg Y.