



Transworld Research Network
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and
Social Networks, 2010: 205-212 ISBN: 978-81-7895-489-9
Editor: Humberto González-Díaz and Cristian Robert Munteanu

12. Study of criminal law networks with Markov-probability centralities

Aliuska Duardo-Sanchez

*Department of Especial Public Law, Financial and Tributary Law Area
Faculty of Law, USC, Santiago de Compostela, 15782, Spain*

Abstract. Graph theory and Complex Network analysis tools are expanding to new potential fields of application at different levels on Information Sciences. For instance, at molecular level we can use them to describe drug-virus action pairs in antiviral medicinal chemistry research. In any case, the applications are far to be restricted to the world of molecules. We can use the same type of graph and complex networks to describe relationships between non-living objects, organisms, or even social actors. Despite of the type of system described we can use numerical parameters of graphs and networks to characterize the structural information of these systems. With these indices, also called Topological Indices (TIs), we can search of Quantitative Structure-Property Relationship (QSPR) models for prediction and discovery of antimicrobial drugs. More in general, we can use TIs to find Quantitative Structure-Property Relationship (QSPR) models for complex social networks. Anyhow, almost all works focus only on the development of QSAR/QSPR models at only one structural level. In this work, we decided to test the potentialities one of the classes of TIs in criminal law networks. For this test we selected the class of TIs called the node absolute

Correspondence/Reprint request: Dr. Aliuska Duardo-Sanchez, Department of Especial Public Law, Financial and Tributary Law Area, Faculty of Law, USC, Santiago de Compostela, 15782, Spain
E-mail: aliuskaduardo@yahoo.es

probabilities $\pi^k(i)$ that can be calculated with the method MARCH-INSIDE based on Markov models. The second model developed is able to discriminate between main and secondary causes in causality criminal case networks with Accuracy = 94.74%. The work opens new directions in the generalization of TIs to develop QSAR/QSPR models for predicting relevant information of systems at different structural levels.

1. Introduction

Graph theory and Complex Network analysis tools are expanding to new potential fields of application of Information Sciences at different levels from molecular to populations, social or technological such as genome networks, protein-protein networks, sexual disease transmission networks, power electric power network or internet [1]. In particular, the case of relationships among social actors, as well as the relationships among actors at different levels of analysis (such as persons and groups) are being subject of intensive investigation [2]. It provides a common approach for all those disciplines involved in social structure study [3-6] susceptible of network depiction. Social structure concept is merely used in sociology and social theory. Although there is not agreement between theorists, it can refer to a specific type of relation between entities or groups also can evolve enduring patterns of behavior and relationship within a society, or social institutions and norms becoming embedded into social systems. For a most complete review of SNA see the in-depth review of Newman M entitled: The Structure and Function of Complex Networks [7]. Anyway, if we take in consideration that a network is a set of items, usually called *nodes*, with connections between them, so called *edges* [8], thus it means the 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 [1]. In fact, SNA is nothing new in social sciences studies, in early 1930s, sociologists already have made a social network to study friendships between school children [9]. Since the important of network approach to social sciences high increased, and its application goes from interrelation between family members [10] to companies business interaction [11, 12] or patterns of sexual contacts [13, 14]. Although the network approach is so pervasive in the social sciences their application in the Law scope is still weak. Networks tools and methodologies might be useful to illustrate the interrelation between the different law types, check the importance of a specific instrument so as the normative hierarchy respect by legislators in order to regulate the most important matter for individuals through law instruments which require the approval from the most representative democratic actors. Also can help to understand laws consequences in society life and its effectiveness or not.

In this sense, using network TIs of application at different levels of organization of matter (molecular, biological, social, economical, or even technological) may be of high interest to develop a general methodology for the search of Quantitative Structure-Property Relationships (QSPR) models. These QSPR models shall connect the structure of the system (drug, protein, microorganism, people, social groups, internet...) with their properties and can be used to predict the behavior of these systems in different situations. Disappointingly, QSPR studies are generally focused on the study of limited properties of small molecules. Anyhow, the applications of TIs in QSPR research are far to cover all the potentialities of TIs and new gateways in molecular, biological, or even social QSPR models are still waiting to be opened. On this line of thinking, our group has introduced a Markov model (MM) method named MARCH-INSIDE: Markovian Chemicals In Silico Design. MARCH-INSIDE generate TIs in the form of matrix invariants such as stochastic entropies, spectral moments, or absolute probabilities for the study of molecular properties. Recently the method has been renamed as MARCH-INSIDE 2.0: Markov Chain Invariants for Network Simultaion & Design, in order to give a more clear idea of the unexplored potentialities. Recent reviews about MACRCH-INSIDE and similar QSPR methods have been published by González-Díaz *et al.* including discussion of multiple applications in different fields [15-19].

In this work, we decided to test the potentialities at different structural levels of one of the classes of TIs calculated by MARCH-INSIDE. For this test we selected the class of TIs called the node absolute probabilities $\pi_k(i)$. The $\pi_k(i)$ values represent the absolute probability of reaching node i after a walk of length k moving from any node in the network. We calculate the $\pi_k(i)$ based on a Markov matrix associated to a graph or network. These TIs differ from other MARCH-INSIDE TIs because they are useful only to describe a node or, if we sum several $\pi_k(i)$ values, we can describe a collection of nodes (atoms, aminoacids, a group of electric plants, a social subgroup...); which form part of larger network systems (molecule, protein, US Electric power system, Society...). It happens because the sum of all $\pi_k(j)$ values for the whole system is always equal to one for any system a do not give structural information. Consequently, the $\pi_k(i)$ values may be consider as local node TIs. We commonly have known this class of TIs as networks nodes Centralities. Several node centralities have been defined before and the software CentiBin calculate some of the more used [20]. However, the definition of new Centralities is an active field of research and new centralities have been recently introduced such as sub-graph centrality introduced by Estrada [21]. Certainly, the $\pi_k(j)$ values were used in the past by our group [22, 23] but ever at the molecular level only and never for no-

molecular problems. In order to both confirm the potentials applications of $\pi_k(j)$ values beyond traditional frontiers, we are going to develop here a new QSPR models. One model is the first QSPR model based on $\pi_k(j)$ values that can be used to predict the probability of crime causation for a single actor (person or cause) in a Criminal law network.

2. Materials and methods

Absolute probability centralities ${}^k C_\pi(j)$ for actions in crime networks

First, we need to construct the crime causality Markov matrix ${}^1\Pi$. This matrix is built up as a square matrix ($n \times n$), where n are all the actions related to the crime including the original actions (causes), the co-actions (secondary causes) and the consequence (crime). The matrix ${}^1\Pi$ contains the transition probabilities (${}^1p_{ij}$) that have the action i to be the cause or at least to be occurred immediately after it in the crime than other action j . The probabilities ${}^1p_{ij}$ may be calculated using the Eq. 28 and 29. δ_j represents the number of actions that occurred immediately after the action i -th. In addition, we use the absolute initial probabilities vector π_0 ; see Eq. 26. This vector lists the absolute initial probabilities ${}^k p_j$ to reach a node n_i from a randomly selected node n_j . Here we consider the initial probability inverse to the dimension (N , number of nodes) of the shp connecting n_i with n_{ii} . Next, we used the theory of Markov chains in order to calculate the criminal causation entropy centrality ${}^k C_\pi(i,ii)$:

$${}^k C_\pi(i,ii) = - \sum_{j \in shp} {}^k C_\pi(j) = - \sum_{j \in shp} {}^k p_j \quad (1)$$

In this equation the values ${}^k p_j$ are the absolute probabilities to reach the nodes moving throughout a walk of length k from node n_i . The sum runs only over the nodes that lie within shp connecting n_i with n_{ii} . The Chapman-Kolmogorov equations were used to calculate the vector π_k containing the ${}^k p_j$ values using the vector π_0 of initial probabilities (${}^0 p_j$) and the matrix ${}^1\Pi$ with the first-step transition probabilities (${}^1 p_{ij}$).

$$\pi_k = \pi_0 \times {}^k \Pi = \pi_0 \times ({}^1 \Pi)^k \quad (2)$$

Data analysis

Using the values of ${}^k C_\pi(i,ii)$, as defined previously, for all pairs of cause(i)-consequence(ii) or the same causality paths we can attempt to

discriminate determinant causes from less important causes in a Crime network. We selected LDA [24] to fit the discriminant function.

$$CC - score = b_0 + a_0 \cdot {}^0C_\pi(i, ii) + a_1 \cdot {}^1C_\pi(i, ii) + \dots + a_k \cdot {}^kC_\pi(i, ii) = b_0 + \sum a_k \cdot {}^kC_\pi(i, ii) \quad (3)$$

In Eq. 37, b_0 and a_k represent the coefficients of the classification function, determined by the least square method implemented in the LDA module of the software STATISTICA 6.0; please see associated book [25]. Forward-stepwise algorithm was used for variable selection [26, 27]. The statistical significance of the LDA model was determined by Fisher's test by examining F and p. All the variables included in the model were standardized in order to bring it into the same scale. Subsequently, a standardized linear discriminant equation that allows to compare their coefficients is obtained [28]. We also inspected the percentage of good classification, cases/variables ratios (ρ parameter), and number of variables to be explored to avoid overfitting or chance correlation [26].

3. Results and discussion

QSPR models for Criminal Causality

One of the reasons people have difficulty in dealing with complex systems is that the linear causal chain way of thinking - A causes B causes C causes D ... etc - breaks down in the presence of feedback and multiple interactions between causal and influence pathways. One could say that complex systems are characterized by networked rather than linear causal relationships. Nevertheless, it is important to be able to reason about complex systems, make inferences about factors that contribute to current and alternative states of complex systems and explore their possible future trajectories, especially if we wish to influence them towards more favorable futures and away from more possibilities that are dangerous. Large scale examples include ecosystems, economic systems, coupled biophysical-socioeconomic systems, integrated supply chains/industrial systems and social systems, but these remarks also apply for example to attempts to understand a physical organism as a complex system. Crime causality is a very important phenomenon in this sense. Different measures of crime causality have been developed before [29]. In this work, we introduced the Markov entropy centrality ${}^kC_\pi(j)$ for a node in a Crime causality network. At the same time, we propose new measures of crime causality calculated as of the sum of all the ${}^kC_\pi(j)$ values of the same order k for all nodes placed in the

shortest path (shp), connecting the original node n_i (possible cause) with the final node n_{ii} (consequence). The model was trained and later validated with and external validation series. The best model found was:

$$CC - score = -35.36 \cdot {}^0C_{\pi}(i, ii) + 112.80 \cdot {}^5C_{\pi}(i, ii) - 25.59 \quad (4)$$

$$n = 47 \quad Rc = 0.85 \quad Chi - sqr = 72.8 \quad p < 0.005$$

The output of the model, CC-score, is a real value variable that scores the possibility of a Crime Cause (CC) to be the main cause of a given crime. This model is able to correctly predict the 94.74% of the main crime causes (CC) out of 47 potential crime causes in 17 crime cases. We also obtained two additional classification functions to discriminate secondary cases of to lower degrees not reported here for reasons of space. The present model also correctly predicts 94.74% of main crime causes in Leave-One-Out (LOO) cross-validation experiments. In **Figure 1**, we illustrate the separation of different crime CN in the canonical space using this LDA model.

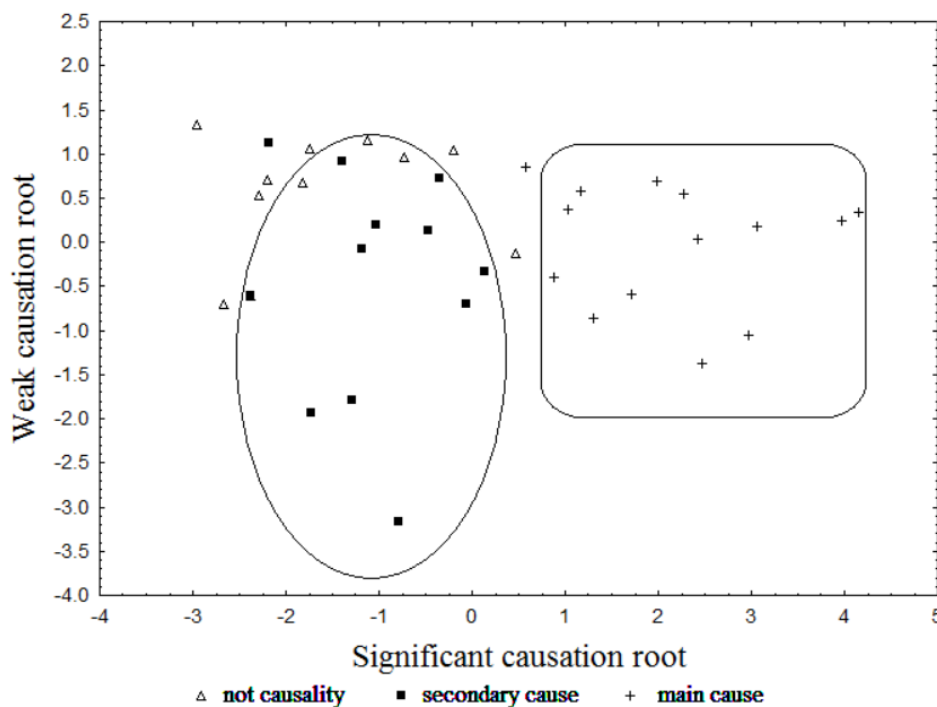


Figure 1. Canonical space representation of different criminal network cases

Acknowledgments

Duardo-Sánchez, A., gratefully acknowledges partial financial support of the Research project (2006/PX 207) from the Department of Especial Public Law, Financial and Tributary Law Area, Faculty of Law, of the University of

Santiago de Compostela in Spain; which was supported by Xunta de Galicia and ESF.

References

1. Bornholdt S, Schuster HG. Handbook of Graphs and Complex Networks: From the Genome to the Internet. Weinheim: WILEY-VCH GmbH & CO. KGa. 2003.
2. Breiger R. The Analysis of Social Networks. In: Hardy M, Bryman A, eds. *Handbook of Data Analysis*. London: Sage Publications 2004:505-26.
3. Abercrombie N, Hill S, Turner BS. Social structure. *The Penguin Dictionary of Sociology*. 4th ed. London: Penguin 2000.
4. Craig C. Social Structure. *Dictionary of the Social Sciences*. Oxford: Oxford University Press 2002.
5. White H, Scott Boorman and Ronald Breiger. . ." Social Structure from Multiple Networks: I Blockmodels of Roles and Positions. *American Journal of Sociology*. 1976;81:730-80.
6. Wellman B, Berkowitz SD. Social Structures: A Network Approach. Cambridge: Cambridge University Press 1988.
7. Newman M. The Structure and Function of Complex Networks. *SIAM Review*. 2003;56:167-256.
8. Newman M. The Structure and Function of Complex Networks. *SIAM Review*. 2003(56):167-256.
9. Moreno JL. Who Shall Survive? New York: Beacon House 1934.
10. Padgett JF, Ansell CKJF. Robust action and the rise of the Medici, 1400-1434. *Amer J Sociol*. 1993;98:259-1319.
11. Mariolis P. Interlocking directorates and control of corporations: The theory of bank control. *Social Sci Quart*. 1975;56:425-39.
12. Mizruchi MS. The American Corporate Network, 1904-1974. Beverly Hills: Sage 1982.
13. 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;38:79-88.
14. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y. The web of human sexual contacts. *Nature*. 2001;411:907-8.
15. Pérez-Montoto LG, Prado-Prado F, Ubeira FM, González-Díaz H. Study of Parasitic Infections, Cancer, and other Diseases with Mass-Spectrometry and Quantitative Proteome-Disease Relationships. *Curr Proteomics*. 2009; 6:in press.
16. González-Díaz H, Prado-Prado F, Pérez-Montoto LG, Duardo-Sánchez A, López-Díaz A. QSAR Models for Proteins of Parasitic Organisms, Plants and Human Guests: Theory, Applications, Legal Protection, Taxes, and Regulatory Issues. *Curr Proteomics*. 2009;6:in press.
17. González-Díaz 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.

18. González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. *Proteomics*. 2008;8:750-78.
19. 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.
20. Junker BH, Koschuetzki D, Schreiber F. Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics*. 2006 Apr 21;7(1):219.
21. Estrada E, Rodriguez-Velazquez JA. Subgraph centrality in complex networks. *Phys Rev E*. 2005 May;71(5 Pt 2):056103.
22. González-Díaz H, Prado-Prado F. Unified QSAR and Network-Based Computational Chemistry Approach to Antimicrobials, Part 1: Multispecies Activity Models for Antifungals. *J Comput Chem*. 2008;29:656-7.
23. González-Díaz H, Sanchez IH, Uriarte E, Santana L. Symmetry considerations in Markovian chemicals 'in silico' design (MARCH-INSIDE) I: central chirality codification, classification of ACE inhibitors and prediction of sigma-receptor antagonist activities. *Comput Biol Chem*. 2003 Jul;27(3):217-27.
24. Estrada E, Molina E. 3D connectivity indices in QSPR/QSAR studies. *J Chem Inf Comput Sci*. 2001 May-Jun;41(3):791-7.
25. Hill T, Lewicki P. *STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining*. Tulsa: StatSoft 2006
26. Van Waterbeemd H. *Chemometric methods in molecular design*. New York: Wiley-VCH 1995.
27. Cruz-Monteagudo M, González-Díaz H, Agüero-Chapin G, Santana L, Borges F, Dominguez ER, et al. Computational chemistry development of a unified free energy Markov model for the distribution of 1300 chemicals to 38 different environmental or biological systems. *J Comput Chem*. 2007 Aug;28(11):1909-23.
28. Kutner MH, Nachtsheim CJ, Neter J, Li W. *Standardized Multiple Regression Model. Applied Linear Statistical Models*. Fifth ed. New York: McGraw Hill 2005:271-7.
29. Devah P. Mark of a Criminal Record. *Am J Soc*. 2003(108):937-75.