Proceedings of the 5th WSEAS Int. Conf. on DISTANCE LEARNING AND WEB ENGINEERING, Corfu, Greece, August 23-25, 2005 (pp234-237)

Graphical Matrices in Chemistry

SONJA NIKOLIC,[a] ANTE MILICEVIC [b] and NENAD TRINAJSTIC [a]

[a]The Rugjer Boskovic Institute, P.O.B. 180, HR-10002 Zagreb, CROATIA
[b]The Institute of Medical Research and Occupational Health, P.O.B. 291,
HR-10001 Zagreb, CROATIA

http://www.irb.hr/

Abstract: - Graphical matrices are presented. Their construction *via* selected sets of subgraphs and the replacement of subgraphs by numbers representing graph invariants are discussed. The last step of the procedure is to apply the method of choice for obtaining the desired double invariant from the graphical matrix in the numerical form. It is also pointed out that many so-called special graph-theoretical matrices from the literature are rooted in the corresponding graphical matrices. Two descriptors, the Wiener-Wiener index and the Hosoya-Wiener index as examples of double invariants have been generated from the appropriate graphical matrices and used in the structure-boiling point and structure-steric energy modeling of octanes. The obtained models possess the statistical parameters that are comparable to those of the best structure-boiling point and structure-steric energy models for octanes in the literature.

Key-Words: - chemical graph theory, double invariants, graphical matrices, Hosoya-Wiener index, molecular descriptors, structure-property modeling, Wiener-Wiener index

1 Introduction

Graphical matrices are matrices whose elements are subgraphs of the graph rather than numbers. Since the elements of these matrices are (sub)graphs, they are called the *graphical matrices* [1]. Thus far a very little work has been reported on these matrices [1,2]. However, many of so-called special matrices [3], such as the Wiener matrices [4-6] and the Hosoya matrices [7], may be regarded as the numerical realizations of the corresponding graphical matrices

The advantage of a graphical matrix lies in the fact that it allows great many possibilities of numerical realizations. In order to obtain a numerical form of a graphical matrix, one needs to select a graph invariant and replace all the graphical elements (subgraphs of some form) by the corresponding numerical values of the selected invariant. In this way, the numerical form of the graphical matrix is established and one can select another or the same type of invariant – this time an invariant of the numerical matrix. Graph invariants generated in this way are *double invariants* [2] in view of the fact that two invariants are used in constructing a given molecular descriptor.

2 Construction of Graphical Matrices and the Numerical Realization

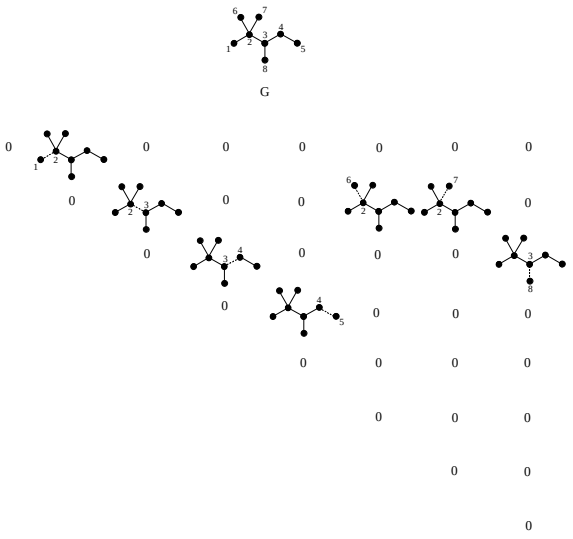
Please, leave two blank lines between successive sections as here. The construction of graphical matrices and the numerical realization consists of the following steps:

1. Representation of the molecule by the corresponding hydrogen-depleted graph;
2. Labelling of the vertices;
3. Construction of the sparse graphical matrix by consecutively removing edges or adjacent vertices from the graph;
4. Construction of the dense graphical matrix by consecutively removing paths of a given length or pairs of end-vertices of paths from the graph;
5. Replacing the elements of the sparse and dense graphical matrices by numerical values of the selected graph invariants to obtain the corresponding numerical matrices, that is, matrices with the numerical elements;
6. Applying a graph invariant of choice to the numerical matrix and getting a double invariant of a graph.

2.1 Example

We illustrate below, as an example, how the *edge-Wiener-Wiener index* for 2,2,3-trimethylpentane is generated. It is so called because the *Wiener index* [8] is used twice – first for the numerical realization of the graphical matrix giving the edge-Wiener matrix and then as the invariant of the numerical matrix. Below we give step 3,5 and 6.

The summation of the matrix-elements in the above matrix-triangle gives the *edge-Wiener-Wiener index* ${}^{e}WW$ of 2,2,3-trimethylpentane (${}^{e}WW$ =274).



G

0	46	0	0	0	0	0	0
	0	19	0	0	46	46	0
		0	29	0	0	0	46
			0	42	0	0	0
				0	0	0	0
					0	0	0
						0	0
							0

3 Testing the Wiener-Wiener Indices

To exemplify the use of the double invariants, we applied the Wiener-Wiener indices as predictor variables in modeling the boiling points, *bp*, and the total steric energies, *SE*, of octane isomers. In selecting octanes as a test set, we followed the advice of Randic [9]. He recommended octanes as a test set because it consists of only 18 isomers that possess sufficient structural variations.

We give for octanes the original Wiener index and all four Wiener-Wiener indices in Table 1. The boiling points [10] (*bp* in ºC) and the total steric energies [11] (*SE* in kJ/mol). of octanes are given in Table 2.

Table 1. Values of the Wiener index (*W*) and the four Wiener-Wiener indices (^{e}WW, ^{p}WW, ^{suv}WW, ^{duv}WW) of octanes

Octanes*	*W*	^{e}WW	^{p}W *W*	^{suv}W *W*	^{duv}W *W*
n-octane	84	252	504	140	378
2-M-heptane	79	262	509	118	398
3-M-heptane	76	268	520	121	416
4-M-heptane	75	270	525	122	423
3-E-hexane	72	276	536	134	441
2,2-MM-hexane	71	270	561	88	438
2,3-MM-hexane	70	274	524	104	449
2,4-MM-hexane	71	274	521	102	443
2,5-MM-hexane	74	270	512	98	423
3,3-MM-hexane	67	274	595	94	464
3,4-MM-hexane	68	276	528	106	462
3,2-EM-pentane	67	278	533	116	469
3,3-EM-pentane	64	276	624	102	483
2,2,3-MMM-pentane	63	274	561	78	485
2,2,4-MMM-pentane	66	274	558	72	467
2,3,3-MMM-pentane	62	274	586	80	491
2,3,4- MMM-pentane	65	276	521	88	477
2,2,3,3-MMMM-butane	58	270	588	54	507

*Abbreviations M and E denote *methyl* and *ethyl*, respectively.

Table 2. The boiling points (*bp* in ºC) and the total steric energies (*SE* in kJ mol^{-1}) of octanes

Octanes*	*bp*†	*SE*‡
n-octane	125.7	0
2-M-heptane	117.6	2.5439
3-M-heptane	118.9	5.0877
4-M-heptane	117.7	5.0877
3-E-hexane	118.5	7.6316
2,2-MM-hexane	106.8	5.8450
2,3-MM-hexane	115.6	11.7905
2,4-MM-hexane	109.4	7.6316
2,5-MM-hexane	109.1	5.0877
3,3-MM-hexane	112.0	11.6901
3,4-MM-hexane	117.7	14.6984
3,2-EM-pentane	115.6	14.6984
3,3-EM-pentane	118.2	17.5351
2,2,3-MMM-pentane	109.8	18.7987
2,2,4-MMM-pentane	99.2	15.6607
2,3,3-MMM-pentane	114.8	22.0999
2,3,4-MMM-pentane	113.5	19.2380
2,2,3,3-MMMM-butane	106.5	27.2002

*Abbreviations M and E denote *methyl* and *ethyl*, respectively
†Taken from G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci*. **39**, 1999, 788-802.
‡Taken from M. Randic, *J. Mol. Struct*. (*Theochem*) **233**, 1991, 45-59.

The structure-property modeling was carried out by the CROMRsel procedure [12-14]. This is a multivariate procedure that has been designed to select the best possible model among the set of models obtained for a given number of descriptors, the criterion being the standard error of estimate. The quality of models is expressed by fitted (descriptive) statistical parameters: the correlation coefficient (R_{fit}), the standard error of estimate (S_{fit}) and the Fisher's test (F). The models are also cross(internally)-validated by a leave-one-out procedure. Statistical parameters for the cross-validated models are symbolized by R_{cv} and S_{cv}, where subscript *cv* denotes the cross-validation.

We considered the structure-boiling point modeling and the structure-steric energy modeling based on the set of indices consisting of the Wiener index and Wiener-Wiener indices. In building these models we considered all possible combinations between indices in Table 1, starting from a single index to up to four indices. We give statistical parameters (R_{fit}, S_{fit}, R_{cv}, S_{cv}, F) for the best structure-boiling point models in Table 3 and for the best structure-steric energy models in Table 4.

Table 3. Statistical parameters R_{fit}, S_{fit}, R_{cv}, S_{cv} and F for the structure-boiling point model based on the Wiener index and the the best structure-boiling point models one, two, three and four Wiener-type indices. Note, N=18 and I=1,…,4

Index	R_{fit}	R_{cv}	$S_{fit}(N\text{-}I\text{-}1)$	$S_{cv}(N\text{-}I\text{-}1)$	F
W	0538	0.374	5.3	5.9	7
${}^{suv}WW$	0.811	0.744	3.7	4.2	31
${}^{suv}WW$, ${}^{duv}WW$	0.842	0.734	3.5	4.3	18
${}^{e}WW$, ${}^{suv}WW$, ${}^{duv}WW$	0.928	0.868	2.5	3.3	29
${}^{e}WW$, ${}^{p}WW$, ${}^{suv}WW$, ${}^{duv}WW$	0.929	0.855	2.6	3.6	21

Table 4. Statistical parameters R_{fit}, S_{fit}, R_{cv}, S_{cv} and F for the structure-steric energy model based on the Wiener index and the best structure-steric energy models with one, two, three and four Wiener-type indices. Note N=18 and I=1,…4

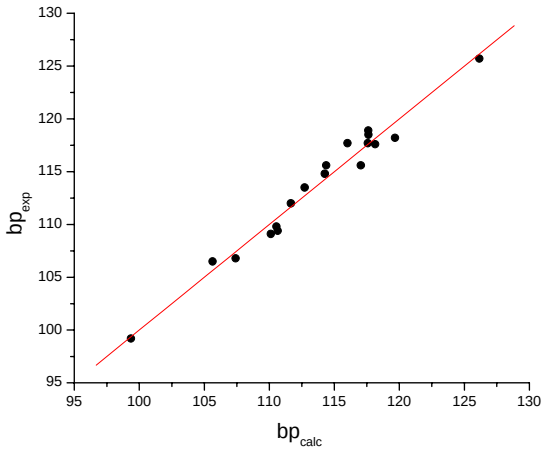
Index	R_{fit}	R_{cv}	$S_{fit}(N\text{-}I\text{-}1)$	$S_{cv}(N\text{-}I\text{-}1)$	F
W	0.958	0.939	2.2	2.6	179
${}^{duv}WW$	0.962	0.946	2.1	2.5	200
${}^{e}WW$, ${}^{duv}WW$	0.982	0.978	1.5	1.6	204
${}^{e}WW$, ${}^{p}WW$, ${}^{duv}WW$	0.990	0.985	1.13	1.4	241
${}^{e}WW$, ${}^{p}WW$, ${}^{suv}WW$, ${}^{duv}WW$	0.992	0.982	1.09	1.6	195

The best structure-boiling point model is based on three descriptors, one being surprisingly the Wiener index:

$$bp = 324\ (\pm 24) - 0.522\ (\pm 0.081)\ W - $$
$$- 1.062\ (\pm 0.084)\ {}^{e}WW + 1.069\ (\pm 0.056)\ {}^{e}ZW$$

R_{fit}=0.987 S_{fit}=1.09 ^{0}C R_{cv}=0.978 S_{cv}=1.40 ^{0}C F=173

Below we give the scatter plot between the experimental and calculated values of octane boiling points for the above model.



Our structure-boiling point model for octanes is comparable with the best models in the literature. For example, the model based on the combination of *twc* (total walk count), $W^{0.25}$ (the fourth root of the Wiener index) and p_3 (the number of paths of length 3) has R_{fit}=0.996 and S_{fit}=0.85 ^{0}C [15].

The structure-steric energy models are fair, but not as good as ones in the literature.

4 Conclusion

This report belongs to our continuous efforts to construct graph invariants of chemical interest and to use them in the structure-property-activity modeling. In the present article, we have discussed an approach for constructing a class of graph-theoretical matrices that are called graphical matrices. They can be used for generating double invariants. So far double invariants found a limited used in QSPR and QSAR modeling. For example, they are only mentioned in passing by in the monumental Todeschini and Consonni's *Handbook of Molecular Descriptors* [16]. In this report, the preliminary results show that the structure-property modeling based on the Wiener-Wiener indices has potential. Therefore, the further exploration of the graphical matrices and their invariants is warranted.

Acknowledgments. - This work was supported by Grant No. 0098034 from the Ministry of Science, Education and Sport of Croatia.

References:

[1] M. Randic, N. Basak and D. Plavsic, *Croat. Chem. Acta* **77,** 2004, 251-257.
[2] M. Randic, D. Plavsic and M Razinger, *MATCH – Comm. Math. Comput. Chem*. **35**, 1997, 243-259.
[3] A. Milicevic, S. Nikolic, N. Trinajstic and D. Janezic, Graph-Theoretical Matrices in Chemistry, *Adv. Quantum Chem.*, in press. Reprints of this paper is available from sonja@irb.hr
[4] M. Randic, *Chem. Phys. Lett*. **211**, 1993, 478-483.
[5] M. Randic, X. Guo, T. Oxley and H. Krishnapryan, *J. Chem. Inf. Comput. Sci*. **33**, 1993, 709-716.
[6] M. Randic, X. Guo, T. Oxley, H. Krishnapryan and L. Naylor, *J. Chem. Inf. Comput. Sci*. **34**, 1994, 361-367.
[7] M. Randic, *Croat. Chem. Acta* **67**, 1994, 415-429.
[8] H. Wiener, *J. Am. Chem. Soc*. **69**, 1947, 19-20.
[9] M. Randic, *Croat. Chem. Acta* **66**, 1993, 289-312.
[10] G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci*. **39,** 1999, 788-802.
[11] M. Randic, *J. Mol. Struct*. (*Theochem*) **233**, 1991, 45-59.
[12] B. Lucic and N. Trinajstic, *J. Chem. Inf. Comput. Sci*. **39**, 1999, 121-132.
[13] T. Pilizota, B. Lucic and N. Trinajstic, *J. Chem. Inf. Comput. Sci*. **44**, 2004, 113-121.
[14] A. Milicevic and S. Nikolic, *Croat. Chem. Acta* **77**, 2004, 97-101.
[15] G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci*. **39,** 1999, 788-802.
[16] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000, pp. 120-121.