On Numerical Characterization of Cyclicity

Tomaž Pisanski,*,† Dejan Plavšić,‡ and Milan Randić§

IMFM, Department of Theoretical Computer Science, University of Ljubljana, Jadranska 19, 1000 Ljubljana, The Republic of Slovenia, The Ruđer Bošković Institute, P.O. Box 1016, HR-10001 Zagreb, The Republic of Croatia, and Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311

Received March 12, 1999

We propose characterizing the cyclicity of molecular graphs by considering their ***D***/***DD*** matrix. Each nondiagonal element of ***D***/***DD*** is a quotient of the corresponding elements of the distance matrix ***D*** and the detour matrix ***DD*** of a graph. In particular, we are using the leading eigenvalue of the ***D***/***DD*** matrix as a descriptor of cyclicity and are investigating for monocyclic graphs C_n how this eigenvalue depends on the number of vertexes n, as n approaches infinity.

1. INTRODUCTION

The distance matrix ***D*** of (molecular) graphs has received considerable attention in mathematical chemistry and has been well studied.[1,2] The elements of the matrix are the distances d_{ij}, where d_{ij} stands for the number of edges on the shortest path between vertexes i and j. The sum of the elements above the main diagonal in ***D*** gives the Wiener number,[3,4] a well-known graph invariant of interest in structure−property correlations. The *detour matrix* ***DD*** of graphs, although suggested in the mathematical literature some time ago,[5] has only recently received attention in mathematical chemistry.[6−10] The elements $(\boldsymbol{DD})_{ij}$ of the detour matrix are the length of the longest path between the vertexes i and j considered. It is interesting to observe that different graphs can have an identical detour matrix (when vertexes are suitably labeled), the first instance of the situation in graph theory that nonisomorphic graphs are represented by an identical matrix.[11]

***D/DD* Matrix.** Construction of novel matrixes and novel graph invariants by using quotients of matrix elements of two different matrixes or two different invariants, respectively, has been introduced in chemical graph theory only relatively recently.[12−21] An example is the matrix whose (i, j) elements are obtained as a quotient between Euclidean and topological (graph theoretical) distance between vertexes i and j. In this case the leading eigenvalue apparently offers a measure of *folding* (bending) of the long chain.[22]

Recently, a new graph matrix, the ***D***/***DD*** matrix, has been introduced.[23] Its diagonal elements are by definition equal to zero, and off-diagonal elements are given as a quotient of the corresponding elements of the distance matrix ***D*** and the detour matrix ***DD***. Note that although the ***DD*** matrix can be identical for nonisomorphic graphs, the distance matrix ***D*** is different for them and consequently the ***D***/***DD*** matrix has to be different for them as well. It has been suggested that the ***D***/***DD*** matrix might offer a novel characterization of cyclic structures.[23] In particular its leading eigenvalue has been put forward as an index of *cyclicity* of a graph.

† University of Ljubljana. E.-mail tomaz.pisanski@fmf.uni-lj.si.
‡ The Ruđer Bošković Institute. E-mail dplavsic@rudjer.irb.hr.
§ Drake University. E-mail milan.randic@drake.edu.

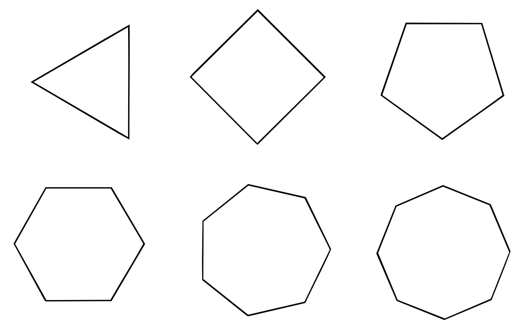

Figure 1. Small monocyclic graphs C_n, where $n = 3, 4, 5, 6, 7,$ and 8.

Cyclicity as well as branching are concepts that have been widely used in chemistry in a qualitative fashion. Attempts to assign to such concepts a numerical magnitude have resulted in different definitions for these quantities that are difficult to define. In a way, the prevailing definition is one that has been found useful, either in the characterization of a model or because it leads to a further development of the field. To illustrate the aforesaid, let us recall of alternative generalizations of the Wiener number, the well-known graph invariant that was defined initially only for trees, to cycle-containing systems.[13−16] Similarly, any definition of cyclicity will be to a degree arbitrary hence the intention is to come up with a definition that will quantify cyclicity and involve as few as possible nonstructural arbitrary choices. To achieve such a goal we should better understand the cyclicity of simple structures first. Having that in mind, we had decided to examine more closely characterization of monocyclic systems. Therefore, we examined how the leading eigenvalue of ***D***/***DD*** varies with n, the size of the cycle graphs, C_n, representing simple monocyclic structures.

2. MONOCYCLIC SYSTEMS

Small monocyclic systems C_n, $n = 3-8$, are illustrated in Figure 1. Cycle graphs, C_n, are vertex- and edge-transitive,[5] leading to adjacency matrixes in which each row can be obtained from the first row by a cyclic permutation. The same applies to the distance and detour matrixes of C_n and consequently to the ***D***/***DD*** matrix of C_n as well. Hence, cycle

10.1021/ci990035m CCC: $19.00
Published on Web 11/06/1999

Table 1. First Row of ***D***/***DD*** Matrixes for Monocycles C_n, Where n is Odd[a]

$n = 3$	0	1/2	1/2								
$n = 5$	0	1/4	2/3	2/3	1/4						
$n = 7$	0	1/6	2/5	3/4	3/4	2/5	1/6				
$n = 9$	0	1/8	2/7	3/6	4/5	4/5	3/6	2/7	1/8		
$n = 11$	0	1/10	2/9	3/8	4/7	5/6	5/6	4/7	3/8	2/9	1/10

[a] Other rows are obtained by cyclic permutation.

Table 2. First Row of ***D***/***DD*** Matrixes for Monocycles C_n, Where n Is Even[a]

$n = 4$	0	1/3	2/2	1/3								
$n = 6$	0	1/5	2/4	3/3	2/4	1/5						
$n = 8$	0	1/7	2/6	3/5	4/4	3/5	2/6	1/7				
$n = 10$	0	1/9	2/8	3/7	4/6	5/5	4/6	3/7	2/8	1/9		
$n = 12$	0	1/11	2/10	3/9	4/8	5/7	6/6	5/7	4/8	3/9	2/10	1/11

[a] Other rows are obtained by cyclic permutation.

Table 3. Row Sums U_n for Smaller Monocycles C_n

n (odd)	U_n	n (even)	U_n
3	1.00000000000	4	1.66666666667
5	1.83333333333	6	2.40000000000
7	2.63333333333	8	3.15238095238
9	3.42142857143	10	3.91269841270
11	4.20396825397	12	4.67705627706

graphs have simple ***D***/***DD*** matrixes. In each row of ***D***/***DD*** of an odd cycle C_{2k+1}, every element appears twice except the diagonal element (Table 1). In case of even cycles C_{2k+2}, in each row of ***D***/***DD*** the largest element and the diagonal element appear once whereas the remaining elements appear twice (Table 2). As it is known from linear algebra,[24] the leading eigenvalue of a symmetric matrix is bounded from above and from below by its largest and smallest row sum, respectively. In the case of vertex-transitive graphs, the two sums are equal; hence for C_n the leading eigenvalue of ***D***/***DD*** equals any row sum. Because the elements of the ***D***/***DD*** matrixes of monocyclic systems have such a simple structure, it is not difficult to write down the explicit expressions for their row sums, U_n:

$$U_n = \begin{cases} 2\displaystyle\sum_{k=1}^{(n-1)/2} \frac{k}{n-k} & \text{if } n \text{ is odd} \\ 1 + 2\displaystyle\sum_{k=1}^{(n/2)-1} \frac{k}{n-k} & \text{if } n \text{ is even} \end{cases} \tag{1}$$

In Table 3 we have listed the values of the row sums U_n of C_n for smaller values of n. It is easy to see that if n increases to infinity, then the row sums U_n also increase indefinitely, even though the increments in each step are getting smaller. In this respect the U_n sums are reminiscent of the harmonic series, which is divergent but at a very slow rate. As it is known, the difference between the harmonic series and the logarithmic function of n leads to the Euler constant γ ($\gamma = 0.5772...$). The sequences built from the nonzero elements of the first row of ***D***/***DD*** matrixes also may be of some interest in mathematics. Consider, for example, the first few nonzero matrix elements of ***D***/***DD*** for the cycle C_{11} (shown in Table 1) and C_{12} (shown in Table 2), respectively:

$$\frac{1}{10}, \frac{2}{9}, \frac{3}{8}, \frac{4}{7}, \frac{5}{6} \qquad \frac{1}{11}, \frac{2}{10}, \frac{3}{9}, \frac{4}{8}, \frac{5}{7}, \frac{6}{6}$$

Table 4. Normalized Row Sums W_n for Monocycles C_n of Increasing Size n, Computed by Mathematica[25]

n (odd)	W_n lower bound	n (even)	W_n upper bound	difference (upper bound − lower bound)
3	0.33333333333	4	0.41666666667	0.08333333334
5	0.36666666667	6	0.40000000000	0.03333333333
7	0.37619047619	8	0.39404761905	0.01785714286
9	0.38015873016	10	0.39126984127	0.01111111111
11	0.38217893218	12	0.38975468976	0.00757575758
19	0.38491122740	20	0.38574280635	0.00263157895
29	0.38570018343	30	0.38684960872	0.00114942529
39	0.38596573794	40	0.38660676359	0.00064102565
99	0.38624334852	100	0.38634435862	0.00010101010
499	0.38629235310	500	0.38620636112	0.00000400802
999	0.38629386012	1000	0.38629486112	0.00000100100
1999	0.38629423600	2000	0.38629448612	0.00000025012
4999	0.38629434111	5000	0.38629438112	0.00000004001
9999	0.38629435612	10000	0.38629436612	0.00000001000
24999	0.38629436032	25000	0.38629436192	0.00000000160

We can observe some resemblance of these sequences with the corresponding harmonic subsequence:

$$\frac{1}{10}, \frac{1}{9}, \frac{1}{8}, \frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1$$

In the next section we consider the asymptotic behavior of U_n.

3. ON THE CONVERGENCE OF NORMALIZED ROW SUMS

The *cyclicity constant* W_n has been defined in this journal, ref 23, as the *normalized row sum* $W_n = U_n/n$. The denominator makes the corresponding series convergent. Table 4, and in particular its lower part, indicates that the normalized row sum converges as $n \to \infty$, and that the convergence is slow. For instance, for $n = 20$ two digits, for $n = 100$ three digits, and for $n = 25\,000$ only eight digits of the limiting value W when $n \to \infty$ are reproduced. Moreover, the normalized row sums W_n for odd n and even n give a lower and an upper bound of W, respectively.

To calculate W with the aforementioned accuracy demands a lot of computations owing to the slow convergence of W_n. Hence, the question arises whether the limit can be calculated analytically to even higher accuracy. The answer is affirmative since we have found out an expression making this possible. Its derivation shall be stated here.

Let us define

$$S_{m,n} = \frac{1}{n}\sum_{k=1}^{m} \frac{k}{n-k} \tag{2}$$

Using $S_{m,n}$ one can write

$$W_n = \begin{cases} \dfrac{2}{n}\displaystyle\sum_{k=1}^{(n-1)/2} \frac{k}{n-k} = 2S_{(n-1)/2,n} & \text{for odd } n \\ \dfrac{1}{n} + \dfrac{2}{n}\displaystyle\sum_{k=1}^{(n/2)-1} \frac{k}{n-k} = \dfrac{1}{n} + 2S_{(n/2)-1,n} & \text{for even } n \end{cases} \tag{3}$$

$S_{m,n}$ can be related to a harmonic series:

$$S_{m,n} = \frac{1}{n}\sum_{k=1}^{m}\frac{k}{n-k} = \frac{1}{n}\sum_{k=1}^{m}\frac{k-n+n}{n-k} = -\frac{m}{n} + \sum_{k=1}^{m}\frac{1}{n-k} = -\frac{m}{n} + \sum_{t=n-1}^{n-m}\frac{1}{t} \quad (4)$$

and by shifting the summation index it can be transformed into

$$S_{m,n} = -\frac{m}{n} + \sum_{t=n-m}^{n-1}\frac{1}{t} = -\frac{m}{n} + \{H(n-1) - H(n-m-1)\} \quad (5)$$

where $H(n)$ is the sum of the initial n terms of the harmonic series:

$$H(n) = 1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n} \quad (6)$$

$S_{m,n}$ has been represented above as a telescopic sum in which the corresponding members of the two harmonic series that diverge will cancel each other. Formally, we can introduce log n function in order to convert the divergent harmonic sequences to convergent sequences:

$$S_{m,n} = -\frac{m}{n} + \{H(n-1) - \log(n-1)\} - \{H(n-m-1) + \log(n-m-1)\} + \{\log(n-1) - \log(n-m-1)\} \quad (7)$$

It is well-known that

$$\lim_{n\to\infty}[(H(n) - \log n)] = \gamma \quad (8)$$

where $\gamma = 0.5772156649...$ is the Euler constant.

For large value of n the cyclicity constant W_n is approximately equal to $2S_{n/2,n}$, being given by

$$S_{n/2,n} = -\frac{1}{2} + \{H(n-1) - \log(n-1)\} - \{H[(n/2)-1] + \log[(n/2)-1]\} + \{\log(n-1) - \log[(n/2)-1]\} \quad (9)$$

Finally, we obtain the sought-after limit W in a closed form:

$$W = \lim_{n\to\infty} W_n = 2\lim_{n\to\infty} S_{n/2,n} = 2(-{}^{1}/_{2} + \gamma - \gamma + \{\log(n-1) - \log[(n/2)-1]\}) = -1 + 2\lim_{n\to\infty}\log\frac{n-1}{(n/2)-1} = 2\log 2 - 1 = 0.38629436112... \quad (10)$$

The advantage of this analytical result for the cyclicity constant of cycle graphs C_n as n tends to infinity is apparent. For instance, in computations for Table 4 we needed 2000 terms to obtain accuracy of about one part per million, because of the slow convergence of the original series. However, if we use the analytical expression it is possible to obtain additional significant digits of W without difficulty. As a mathematical curiosity we note that the Euler constant γ does not appear in the expression for W contrary to what one would expect.

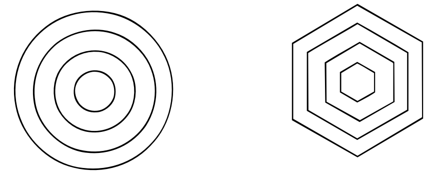

Figure 2. Continuous and discretized circles of increasing radius.

Table 5. Alternative Measures of Smoothness of Discretized n-Circles

n	$W_n - W$	W_n/n
3	−0.05296102779	0.11111111111
4	0.03037230555	0.10416666667
5	−0.01962769445	0.07333333333
6	0.01370563888	0.06666666667
7	−0.01010388493	0.05374149660
8	0.00775325838	0.04925595238
9	−0.00613563096	0.04223985891
10	0.00497548015	0.03912698413
11	−0.00411542894	0.03474353929
12	0.00346032864	0.03247955748
19	−0.00138313372	0.02025848565
20	0.00124844523	0.01937714032
29	−0.00059417769	0.01330000633
30	0.00055524760	0.01289498696
39	−0.00032862318	0.00989655738
40	0.00031240247	0.00966516909
99	−0.00005101260	0.00390144796
100	0.00004999750	0.00386344359
499	−0.00000200802	0.00077413297
500	0.00000200000	0.00077259272
999	−0.00000050100	0.00038668054
1000	0.00000050000	0.00038629486
1999	−0.00000012512	0.00019324374
2000	0.00000012500	0.00019314724
4999	−0.00000002001	0.00007727432
5000	0.00000002000	0.00007725888
9999	−0.00000000500	0.00003863330
10000	0.00000000500	0.00003862944
24999	−0.00000000080	0.00001545239
25000	0.00000000080	0.00001545177

4. ON THE INTERPRETATION OF THE CYCLICITY MEASURE W_N

In Table 4 we listed the difference between W_n for adjacent n values. As n increases, the difference is decreasing and tends to zero. To arrive at an interpretation of W_n we have to consider what other structural elements of C_n approach zero as n tends to infinity. The curvature, which in the case of a circle is given by $1/R$, tends to zero as R increases, i.e., as finite segments of circle approach line. W_n is independent of the geometrical scale; thus it cannot have relation to the curvature. We may, however, consider a discrete analogue of curvature defined by the angle θ_n between the sides of a regular n-gon. In contrast to the concept of curvature in geometry, which is scale-dependent (see Figure 2), now curvature is scale-independent. Thus, for example, curvature of all hexagons of Figure 2 is constant, while that of concentric circles decreases as R increases.

Discrete curvature is a measure of departure of an n-gon from a circle. Clearly, as n increases the difference between n-gon and circle decreases (which has historically been the basis for the early calculation of π). Thus we can take the difference $W_n - W$ (given in Table 5) as a measure of "smoothness" of discretized circles.

An alternative approach is to consider instead of W_n the quantity W_n/n (see Table 5). This quantity has an advantage over W_n and the difference $W_n - W$ in that it does not alternates with even/odd parity changes, but as we see from Table 5 it monotonically decreases as $n \rightarrow \infty$. Clearly W_n/n can be taken as a measure of the smoothness of discretized circles.

The above interpretation answers a number of questions that could be raised when considering numerical characterization of cyclic structures is considered. It is clear now why the convergence of W_n is so important. It is not merely a matter of computation, but the approach offers a basis for measuring the smoothness of discrete curves. We hope that more light will be brought to characterization of cyclicity by extending the present work to polycyclic systems.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Technology of the Republic of Slovenia through Grant J1-8901-0101-98, the Ministry of Science and Technology of the Republic of Croatia, and Croatian–Slovenian project Discrete Mathematics in Chemistry.

REFERENCES AND NOTES

(1) Buckley, F.; Harary, F. *Distance in Graphs*; Addison-Wesley: Reading, MA, 1990.
(2) Mihalić, Z.; Veljan, D.; Amić, D.; Nikolić, S.; Plavšić, D.; Trinajstić, N. The Distance Matrix in Chemistry. *J. Math. Chem.* **1992**, *11*, 223–258.
(3) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
(4) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
(5) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.
(6) Lukovits, I. The Detour Index. *Croat. Chem. Acta* **1996**, *69*, 873–882.
(7) Nikolić, S.; Trinajstić, N.; Jurić, A.; Mihalić, Z. The Detour Matrix and the Detour Index of Weighted Graphs. *Croat. Chem. Acta* **1996**, *69*, 1577–1591.
(8) Lukovits, I.; Razinger, M. On Calculation of the Detour Index. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 283–286.
(9) Trinajstić, N.; Nikolić, S.; Lučić, B.; Amić, D.; Mihalić, Z. The Detour Matrix in Chemistry. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 631–638.
(10) Trinajstić, N.; Nikolić, S.; Mihalić, Z. On Computing the Molecular Detour Matrix. *Int. J. Quantum Chem.* **1997**, *65*, 415–419.
(11) Randić, M.; DeAlba, L. M.; Harris, F. E. Graphs with the Same Detour Matrix. *Croat. Chem. Acta* **1998**, *71*, 53–68.
(12) Randić, M. Search for Optimal Molecular Descriptors. *Croat. Chem. Acta* **1991**, *64*, 43–54.
(13) Mohar, B.; Babić, D.; Trinajstić, N. A Novel Definition of the Wiener Index for Trees. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 153–154.
(14) Randić, M. Novel Molecular Descriptor for Structure–Property Studies. *Chem. Phys. Lett.* **1993**, *211*, 478–483.
(15) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H. Wiener Matrix: Source of Novel Graph Invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709–716.
(16) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener Matrix Invariants. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361–367.
(17) Randić, M.; Mihalić, Z.; Nikolić, S.; Trinajstić, N. Graphical Bond Orders: Novel Structural Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 403–409.
(18) Plavšić, D.; Šoškić, M.; Landeka, I.; Trinajstić, N. On the Relation between the P′/P Index and the Wiener Number. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1123–1126.
(19) Plavšić, D.; Šoškić, M.; Lerš, N. On the Calculation of the Molecular Descriptor χ'/χ. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 889–892.
(20) Plavšić, D. On the Definition and Calculation of the Molecular Descriptor R′/R. *Chem. Phys. Lett.* **1999**, *304*, 111–116.
(21) Plavšić, D.; Lerš, N.; Sertić-Bionda, K. On the Relation between W′/W Index, Hyper-Wiener Index, and Wiener Number. *J. Chem. Inf. Comput. Sci.*, in press.
(22) Randić, M.; Kleiner, A. F.; DeAlba, L. M. Distance/Distance Matrix. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277–286.
(23) Randić, M. On Characterization of Cyclic Structures. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1063–1071.
(24) Roberts, F. S. *Applied Combinatorics*; Prentice Hall: Englewood Cliffs, NJ, 1984.
(25) Wolfram, S. *Mathematica, A System of Doing Mathematics by Computer*; 2nd ed.; Addison-Wesley: Reading, MA, 1991.

CI990035M