STUDIA UBB. CHEMIA, LV, 4, 2010

# CENTRIC CONNECTIVITY INDEX

**MIRCEA V. DIUDEA**[*]

**ABSTRACT.** Relative centricity *RC* values of vertices/atoms are calculated within the Distance and Cluj-Distance criteria. The vertex *RC* distribution in a molecular graph provides atom equivalence classes, useful in interpretation of NMR spectra. Timed by vertex valences, *RC* provides a new index, called Centric Connectivity *CC*, which can be useful in topological characterization of graphs and in QSAR/QSPR studies.

***Keywords**: graph theory, Cluj index, relative centricity, centric connectivity index*

## INTRODUCTION

Let $G = (V, E)$ be a connected graph, with no multiple bonds and loops. $V$ is the set of vertices and $E$ is the set of edges in $G$; $v = |V(G)|$ and $e = |E(G)|$ are their cardinalities.

A *walk* $w$ is an alternating string of vertices and edges: $w_{1,n} = (v_1, e_1, v_2, e_2, ..., v_{n-1}, e_m, v_n)$, with the property that any subsequent pair of vertices represent an edge: $(v_{i-1}, v_i) \in E(G)$. Revisiting of vertices and edges is allowed [1-6].

The *length* of a walk, $l(w_{1,n}) = |E(w_{1,n})|$ equals the number of its traversed edges. In the above relation $E(w_{1,n})$ is the edge set of the walk $w_{1,n}$ . The walk is *closed* if $v_1 = v_n$ and is *open* otherwise [3,5].

A *path* $p$ is a walk having all its vertices and edges distinct: $v_i \neq v_j$, $(v_{i-1}, v_i) \neq (v_{j-1}, v_j)$ for any $1 \leq i < j \leq n$. As a consequence, revisiting of vertices and edges, as well as branching, is prohibited. The *length* of a path is $l(p_{1,n}) = |E(p_{1,n})| = |V(p_{1,n})| - 1$, with $V(p_{1,n})$ being the vertex set of the path $p_{1,n}$. A closed path is a *cycle* ( i.e., *circuit*).

A path is *Hamiltonian* if all the vertices in $G$ are visited at most once: $n = |V(G)|$. If such a path is closed, then it is a *Hamiltonian circuit.*

The *distance* $d_{ij}$ is the length of a *shortest* path joining vertices $v_i$ and $v_j$: $d_{ij} = \min l(p_{ij})$; otherwise $d_{ij} = \infty$. The *set of all distances* (i.e., geodesics) in $G$ is denoted by **DI**($G$).

The *detour* $\delta_{ij}$ is the length of a *longest* path between vertices $v_i$ and $v_j$: $\delta_{ij} = \max l(p_{ij})$; otherwise $\delta_{ij} = \infty$. The *set of all detours* in $G$ is denoted by **DE**($G$).

The square arrays that collect the distances and detours, in $G$ are called the *Distance* **DI** and *Detour* **DE** matrix, respectively [3,5]:

$$[\mathbf{DI}(G)]_{i,j} = \begin{cases} \min l(p_{i,j}), \text{if } i \neq j \\ 0 \text{ if } i=j \end{cases} \quad (1)$$

---

[*] *Faculty of Chemistry and Chemical Engineering,Babes-Bolyai University, 3400 Cluj, Romania*

$$|\mathbf{DE}(G)|_{i,j}=\begin{cases}\max l(p_{i,j}), if & i\neq j\\ 0 \quad if & i=j\end{cases} \tag{2}$$

In words, these matrices collect the number of edges separating the vertices $i$ and $j$ on the shortest and longest path $p_{i,j}$, respectively. The half sum of entries in the Distance and Detour matrices provide the well-known Wiener index $W$ [7] and its analogue, the detour number $w$ [8,9].

The *Cluj fragments* are sets of vertices obeying the relation [3,5,10-13]:

$$CJ_{i,j,p}=\left\{v\middle| v\in V(G); D_{(G-p)}(i,v)<D_{(G-p)}(j,v)\right\} \tag{3}$$

The entries in the Cluj matrix **UCJ** are taken, by definition, as the maximum cardinality among all such fragments:

$$[\mathbf{UCJ}]_{i,j}=\max_{p}\left|CJ_{i,j,p}\right| \tag{4}$$

It is because, in graphs containing rings, more than one path can join the pair ($i$, $j$), thus resulting more than one fragment related to $i$ (with respect to $j$ and path $p$).

The Cluj matrix is defined by using either *distances* or *detour*s [14]: when path $p$ belongs to the set of distances DI($G$), the suffix DI is added to the name of matrix, as in UCJDI. When path p belongs to the set of detours DE($G$), the suffix is DE.

Two graphs are called *isomorphic*, G ≈ G', if there exists a mapping $f: V \rightarrow V'$ that preserves adjacency (*i.e.*, if $(i,j)\in$ E(G), then $(f(i), f(j))\in$ E'(G')). The function $f$ provides a one-to-one correspondence between the vertices of the two sets. The isomorphism of $G$ with itself is called an automorphism. It is demonstrated that all the automorphisms of $G$ form a group, *Aut*(G) [3,5].

The symmetry of a graph is often called a topological symmetry; it is defined in terms of *connectivity*, as a constitutive principle of molecules and expresses equivalence relationships among elements of the graph: vertices, bonds, faces or larger subgraphs. The topological symmetry does not fully determine molecular geometry and it does not need to be the same as (*i.e.*, isomorphic to) the molecular point group symmetry. However, it represents the maximal symmetry which the geometrical realization of a given topological structure may posses [15-17].

Given a graph $G=(V,E)$ and a group Aut($G$), two vertices, $i, j\in$V are called *equivalent* if there is a group element, aut($v_i$)∈Aut($G$), such that $j$ aut($v_i$) $i$. The set of all vertices $j$ (obeying the *equivalence relation*) is called the $i$'s class of equivalence. Two vertices $i$ and $j$, showing the same vertex invariant $In_i=In_j$ belong to the same *invariant class IC*. The process of vertex partitioning in *IC*-s leads to $m$ classes, with $v_1$, $v_2$,...$v_m$ vertices in each class. Note that invariant-based partitioning may differ from the orbits of automorphism since no vertex invariant is known so far to discriminate two non-equivalent vertices in any graph [3,5].

In the chemical field, the isomorphism search could answer to the question if two molecular graphs represent or not one and the same chemical compound. Two isomorphic graphs will show the same topological indices, so that they cannot be distinguished by topological descriptors.

## CENTRIC CONNECTIVITY CC INDEX

In studies on the centrality/centricity of graphs, Bonchev *et al.* [18,19] have proposed the distance-based criteria *1D-3D* as follows:

1D*:* minimum vertex eccentricity: min $ecc_i$
2D*:* minimum vertex distance sum*:* min $DIS_i$
3D*:* minimum number of occurrence of the largest distance*:*
min $[\mathbf{LM}, \mathbf{ShM}]_{i,j\,\max}$

When applied hierarchically, the above criteria lead to the center(s) of a graph.

Our older centrality index *C*(**LM**, **ShM**) is a function also giving the graph center(s), used alone or within the MOLORD algorithm [20]. In the above, **LM**, **ShM** denote the layer matrix and the shell matrix (of a given square info-matrix M), defined as follows [21-23].

The entries in the layer matrix (of vertex property) **LM**, are

$$[\mathbf{LM}]_{i,k} = \sum_{v|d_{i,v}=k} p_v \tag{5}$$

Layer matrix is a collection of the above defined entries:

$$\mathbf{LM}(G) = \left\{ [\mathbf{LM}]_{i,k};\, i \in V(G);\, k \in [0,1,..,d(G)] \right\} \tag{6}$$

with *d*(*G*) being the diameter of the graph (*i.e.*, the largest distance in *G*). Any atomic/vertex property can be considered as $p_i$. More over, any square matrix **M** can be taken as *info matrix*, *i.e.*, the matrix supplying local/vertex properties as row sum *RS*, column sum *CS.* The zero column is just the column of vertex properties $[\mathbf{LM}]_{i,0} = p_i$. When the vertex property is 1 (*i.e.*, the counting property), the **LM** matrix will be **LC** (the Layer matrix of Counting).

Define the entries in the shell matrix **ShM** (of pair vertex property) as [23]

$$[\mathbf{ShM}]_{i,k} = \sum_{v|d_{i,v}=k} [\mathbf{M}]_{i,v} \tag{7}$$

The shell matrix is a collection of the above defined entries:

$$\mathbf{ShM}(G) = \left\{ [\mathbf{ShM}]_{i,k};\, i \in \mathrm{V}(G);\, k \in [0,1,..,d(G)] \right\} \tag{8}$$

A shell matrix **ShM**(*G*), will partition the entries of the square matrix according to the vertex (distance) partitions in the graph. It represents a true decomposition of the property collected by the info square matrix according to the contributions brought by pair vertices pertaining to shells located at distance *k* around each vertex. The zero column entries $[\mathbf{ShM}]_{i,0}$ are just the diagonal entries in the info matrix.

In this paper, the distance-based functions, expressing the topology related to the center of the graph, are as follows:

$$EP(i) = \sum\nolimits_k P(i)_k \cdot k^n;\, k = 1,2,...d(G);\, n = 1,2,.. \tag{9}$$

$$P(i)_k = [\mathbf{LM}, \mathbf{ShM}]_{i,k} \tag{10}$$

$$C(i) = (EP(i))^{-1} \tag{11}$$

$$RC(i) = C(i) / C(i)_{\max} = EP(i)_{\min} / EP(i) \tag{12}$$

$$RC(G) = \sum\nolimits_i RC(i) \tag{13}$$

$$CC(i) = RC(i) \cdot d(i) \tag{14}$$

$$CC(G) = \sum\nolimits_i CC(i) \tag{15}$$

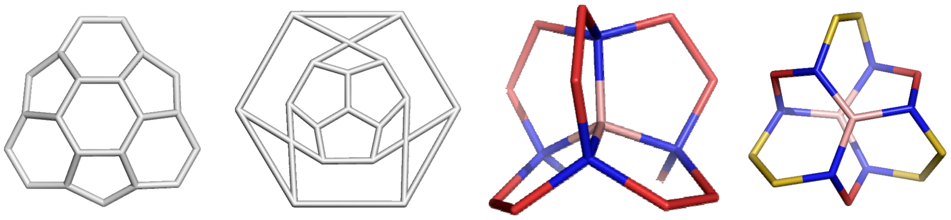
The distance-extension of the property *P*(*i*) (collected in **LM** or **ShM**, (10)) is made by a variable power function, depending of the info matrix M, to ensure the separation of the resulting values, of which meaning is that of an eccentric property *EP*(i) (9). There is a clear difference between *EP*(i) and the eccentricity $\varepsilon(i)$ (counting the largest topological distance from *i* to any other vertex in *G*), used in the construction of "Eccentric Connectivity index" [24]. The vertex centricity *C*(i) (11) is then calculated in the sense of the Bonchev's 1D-3D criteria, by virtue of the involved **LM, ShM** matrices.

The relative centricity (or centrality) *RC*(i) (12) accounts for the deviation to the maximum centrality, equaling 1 in case of vertices being centers of the graph. The global value RC(G) (13) is useful in characterizing the distribution of the centricity function (11), particularly when normalized by the number of vertices of G.

Finally, the centric connectivity CC index (14,15) is hoped to be useful in QSAR/QSPR studies, their values being of the same order of magnitude as the number of vertices/atoms in the molecular graph. Relation (14) can be generalized by changing *d*(i) by the "remote" degree [5,25] or by degrees of "extended connectivity" [26-30].

Tables 1 and 2 exemplify the above formulas for the molecular graphs illustrated in Figure 1. The sum in the EP(i) column gives twice the Wiener index [7]. Note that $G_2$ (Table 1) is a self-centered graph [31], of which all vertices are centers of the graph, as ranked by the RC(i) column. Also note that $G_2$ is a full Hamiltonian detour graph [14]; this means that all its detours are Hamiltonian path, visiting once all the vertices of the graph.

The invariant classes of equivalence *IC*-s are given at the bottom of tables, by their population Pop (no. of vertices in each class). *IC*-s are important in NMR spectra interpretation.



$G_1$ $G_2$ $G_3$ $G_4$

**Figure 1.** $G_1$ (v=21); $G_2$ (v=16); $G_3$ (v=17); $G_4$ (v=17)

**Table 1.** Vertex eccentric property EP(i), relative centricity RC(i) and centric connectivity CC(i) values, calculated on **LM**=**LC** matrix (k^1), unless otherwise specified; no. of "invariant classes" IC-s, by their population Pop (no. of vertices) for $G_1$ and $G_2$.

| | $G_1$ | | | | $G_2$ (self-centered graph) | | | |
|---|---|---|---|---|---|---|---|---|
| # | EP(i) | RC(i) | CC(i) | d(i) | EP(i) | RC(i) | CC(i) | CC(i); (**ShCJDI;** k^2) |
| 1 | 70 | 0.7714 | 1.5429 | 2 | 33 | 1 | 3 | 2.8385 |
| 2 | 70 | 0.7714 | 1.5429 | 2 | 33 | 1 | 3 | 2.9611 |
| 3 | 70 | 0.7714 | 1.5429 | 2 | 33 | 1 | 3 | 2.9358 |
| 4 | 54 | 1 | 3 | 3 | 33 | 1 | 3 | 2.9358 |
| 5 | 64 | 0.8438 | 2.5313 | 3 | 33 | 1 | 3 | 3 |
| 6 | 74 | 0.7297 | 1.4595 | 2 | 33 | 1 | 3 | 2.9611 |
| 7 | 64 | 0.8438 | 2.5313 | 3 | 33 | 1 | 3 | 2.9358 |
| 8 | 74 | 0.7297 | 1.4595 | 2 | 33 | 1 | 3 | 3 |
| 9 | 64 | 0.8438 | 2.5313 | 3 | 33 | 1 | 3 | 2.9611 |
| 10 | 54 | 1 | 3 | 3 | 33 | 1 | 3 | 2.9358 |
| 11 | 64 | 0.8438 | 2.5313 | 3 | 33 | 1 | 3 | 3 |
| 12 | 54 | 1 | 3 | 3 | 33 | 1 | 3 | 3 |
| 13 | 64 | 0.8438 | 2.5313 | 3 | 33 | 1 | 3 | 2.9358 |
| 14 | 74 | 0.7297 | 1.4595 | 2 | 33 | 1 | 3 | 2.9358 |
| 15 | 64 | 0.8438 | 2.5313 | 3 | 33 | 1 | 3 | 3 |
| 16 | 74 | 0.7297 | 1.4595 | 2 | 33 | 1 | 3 | 3 |
| 17 | 54 | 1 | 3 | 3 | 528=2W | 16 | 48=CC | 47.3364=CC |
| 18 | 54 | 1 | 3 | 3 | | | | |
| 19 | 74 | 0.7297 | 1.4595 | 2 | | | | |
| 20 | 54 | 1 | 3 | 3 | | | | |
| 21 | 74 | 0.7297 | 1.4595 | 2 | | | | |
| Sum | 1362=2W | 17.7552 | 46.5728=CC | | | | | |
| Pop | | | $3,(6)_3$ | | | | | $1,3,(6)_2$ |

**Table 2.** Vertex eccentric property EP(i), relative centricity RC(i) and centric connectivity CC(i) values, calculated on **ShM**=**ShUCJDI** matrix (k^3); no. of "invariant classes" *IC*-s, by their population Pop (no. of vertices) for $G_3$ and $G_4$;

| # | $G_3$ | | | | $G_4$ | | | |
|---|---|---|---|---|---|---|---|---|
| | EP(i) | RC(i) | CC(i) | d(i) | EP(i) | RC(i) | CC(i) | d(i) |
| 1 | 2478 | 0.5585 | 1.1170 | 2 | 2483 | 0.7962 | 2.3886 | 3 |
| 2 | 2478 | 0.5585 | 1.1170 | 2 | 2483 | 0.7962 | 2.3886 | 3 |
| 3 | 2044 | 0.6771 | 2.7084 | 4 | 2483 | 0.7962 | 2.3886 | 3 |
| 4 | 1384 | 1 | 4 | 4 | 1977 | 1 | 3 | 3 |
| 5 | 2044 | 0.6771 | 2.7084 | 4 | 3322 | 0.5951 | 1.1902 | 2 |
| 6 | 2478 | 0.5585 | 1.1170 | 2 | 2982 | 0.6630 | 1.3260 | 2 |
| 7 | 2478 | 0.5585 | 1.1170 | 2 | 2483 | 0.7962 | 2.3886 | 3 |
| 8 | 2044 | 0.6771 | 2.7084 | 4 | 2982 | 0.6630 | 1.3260 | 2 |
| 9 | 2478 | 0.5585 | 1.1170 | 2 | 2483 | 0.7962 | 2.3886 | 3 |
| 10 | 2478 | 0.5585 | 1.1170 | 2 | 3322 | 0.5951 | 1.1902 | 2 |
| 11 | 2044 | 0.6771 | 2.7084 | 4 | 2483 | 0.7962 | 2.3886 | 3 |
| 12 | 2478 | 0.5585 | 1.1170 | 2 | 1977 | 1 | 3 | 3 |
| 13 | 2478 | 0.5585 | 1.1170 | 2 | 3322 | 0.5951 | 1.1902 | 2 |
| 14 | 2478 | 0.5585 | 1.1170 | 2 | 2982 | 0.6630 | 1.3260 | 2 |
| 15 | 2478 | 0.5585 | 1.1170 | 2 | 3322 | 0.5951 | 1.1902 | 2 |
| 16 | 2478 | 0.5585 | 1.1170 | 2 | 3322 | 0.5951 | 1.1902 | 2 |
| 17 | 2478 | 0.5585 | 1.1170 | 2 | 3322 | 0.5951 | 1.1902 | 2 |
| Sum | 39296 | 10.4106 | 28.2380 | | 47730 | 12.3370 | 31.4512 | |
| Pop | | | 1,4,12 | | | | $2,3,(6)_2$ | |

## CONCLUSIONS

The relative centricity *RC* of vertices in a molecular graph were calculated within the Distance and Cluj-Distance criteria. The vertex *RC* distribution in a molecular graph gives information on the equivalence classes (as vertex invariant classes) of atoms, useful in interpretation of NMR spectra. Timed by vertex valences, *RC* provides an index, called Centric Connectivity *CC*. This index represents a new descriptor, which can be useful in topological characterization of graphs and in QSAR/QSPR studies. By definition, there is a clear difference between the Centric Connectivity index and the older Eccentric Connectivity index or its newer versions [32-34].

## REFERENCES

1. N. Trinajstić, *Chemical Graph Theory,* CRC Press: Boca Raton, FL, **1983**.
2. F. Harary, *Graph Theory*, Addison-Wesley, Reading, M.A., **1969**.
3. M.V. Diudea, I. Gutman, L. Janschi, *Molecular Topology*, Nova Science, Huntington, N.Y., **2001**.
4. M.V. Diudea, O. Ivanciuc, *Molecular Topology*, Comprex, Cluj, **1995** (in Romanian).
5. M.V. Diudea, *Nanomolecules and Nanostructures, Polynomials and Indices*, MCM, No. 10, Univ. Kragujevac and Fac. Sci. Kragujevac, Serbia, **2010**.
6. M.V. Diudea, M.S. Florescu, P.V. Khadikar, *Molecular Topology and Its Applications*, EFICON, Bucharest, **2006**.
7. H. Wiener, *J. Am. Chem. Soc.,* **1947**, *69*, 17.
8. I. Lukovits, *Croat. Chem. Acta,* **1996**, *69*, 873.
9. I. Lukovits and M. Razinger, *J. Chem. Inf. Comput. Sci*., **1997**, *37*, 283.
10. M.V. Diudea, *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 300.
11. M.V. Diudea, *MATCH, Commun. Math. Comput. Chem*., **1997**, *35*, 169.
12. M.V. Diudea, I. Gutman, *Croat. Chem. Acta,* **1998**, *71*, 21.
13. M.V. Diudea, *Croat. Chem. Acta,* **1999**, *72*, 835.
14. M.V. Diudea, B. Parv, I. Gutman, *J. Chem. Inf. Comput. Sci.,* **1997**, *37*, 1101.
15. C.Y. Hu, L. Xu, *Anal. Chim. Acta,* **1994**, *295*, 127.
16. G.S. Ezra, *Symmetry Properties of Molecules*, Lecture Notes in Chemistry 28, Springer, **1982**.
17. M. Razinger, K. Balasubramanian, M.E. Munk, *J. Chem. Inf. Comput. Sci.*, **1993**, *33*, 197.
18. D. Bonchev, A.T. Balaban, M. Randić, *Int. J. Quantum Chem*., **1981**, *19*, 61.
19. D. Bonchev, O. Mekenyan, A.T. Balaban, *J. Chem. Inf. Comput. Sci*., **1989**, *29*, 91.
20. M.V. Diudea, D. Horvath, D. Bonchev, *Croat. Chem. Acta*, **1995**, *68*, 131.
21. M.V. Diudea, *J. Chem. Inf. Comput. Sci.,* **1994**, *34*, 1064.
22. M.V. Diudea, M. Topan, A. Graovac, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 1071.
23. M.V. Diudea, O. Ursu, *Indian J. Chem*., 42*A*, **2003**, 1283.
24. V. Sharma, R. Goswami, A.K. Madan, *J. Chem. Inf. Comput. Sci*., **1997**, *37*, 273.
25. P.E. John, M.V. Diudea, *Carpath. J. Math.*, **2004**, *20* (2), 235.
26. A.T. Balaban, O. Mekenyan, D. Bonchev, *J. Comput. Chem.*, **1985**, *6*, 538.
27. A.T. Balaban, O. Mekenyan, D. Bonchev, *J. Comput. Chem.,* **1985**, *6*, 552.
28. O. Mekenyan, A.T. Balaban, D. Bonchev, *J. Magn. Reson.,* **1985**, *63*, 1.
29. A.T. Balaban, I. Moțoc, D. Bonchev, O. Mekenyan, *Top. Curr. Chem*., **1993**, *114*, 21.
30. H Morgan, *J. Chem. Doc.*, **1965**, *5*, 107.
31. F. Buckley, *Ann. New York Acad. Sci*., **1989**, *576*, 71.
32. A. Ilić, I. Gutman, *MATCH Commun. Math. Comput. Chem*., **2011**, *65*, 731.
33. A. Ilić, in: I. Gutman, B. Furtula, “Novel Molecular Structure Descriptors -Theory and Applications II”, MCM 9, University of Kragujevac, **2010**.
34. G. Yu, L. Feng, A. Ilić, *J. Math. Anal. Appl*., **2011**, *375*, 934.