Topological Modelling of Chemical Graphs Using Wiener and Zagreb Indices

***Haider Abdulsada*[1*], *Faik Mayah*[2*]**

[1] Department of Mathematics, College of Education for Pure Sciences, Wasit University, Iraq.
[2]College of Science, Wasit University, Iraq.

Abstract. This research delves into chemical topology, am method that uses graph theory to represents molecules. In this framework, atoms re depicted as points (vertices) and the bonds connecting them as lines (edges). We examined the topological characteristics of several hydrocarbons by translating each molecule into graph d then analysing its specific graph. The values of two different topological indices, one related to the distances of the vertices of the graph and the other related to the degrees of the vertices of the graph, have been calculated which are called Wiener and the first Zagreb indices. These indices can further be analyzed to determine the chemical properties of the compounds. The methodology for this research has been to utilize the vertex distance matrices of the graphs of the compounds. The sub-graphs and the derived graphs of the compounds have also been calculated. Further, the induced topologies are obtained and then find the relationship between these indices and size of topology and getting a strong correlation. The results of the research have shown that the Wiener Index is related to the structural organization of the compounds, and the first Zagreb Index is related to the representation of the compounds.

Keywords: Graph; Topological indices; Pearson Correlation; Semi Linear Regression**;** Topological space.

1 Introduction

Chemical applications are among the most important and rapidly developing research fields in recent times. "In this study, we address the role of Graph Theory in chemistry, emphasizing the robust framework it provides for accurately representing molecular structures in a way that. In this framework, compounds or molecules are represented graphically, essentially transforming complex structures into simple mathematical representations that are easily analysed. Within this framework, atoms are visualized as points (vertices), and the bonds between them as lines (edges) [1].

* Corresponding author: std.2024205.h.abdull@uowasit.edu.iq; fmayah@uowasit.edu.iq

There are many applications of graph theory in chemistry, one of which is topological indicators, which assign a numerical value to a molecule. Furthermore, the importance of indicators in modern chemistry, or as they are sometimes referred to in the literature, the genetic fingerprint of a molecule, allows for the interpretation of physical behavior and chemical activity through numerical values known as topological indicators. Some relationships, such as QSPR and QSAR, are used to predict properties like stability, boiling points, and polarity [2].

The Wiener graph index, defined as the sum of the semi-semiconductor distances between each pair of vertices (atoms), was developed by chemist Harold Wiener in 1947. It can link the graph structure of a molecule to its physical properties or predict the properties of alkanes. There is also a relationship between this index and the boiling point of a chemical compound [3].

The Zagreb I index is defined as the sum of the squares of the scores of all nodes (atoms), i.e., it depends on the score of each vertex in the graph. It has a high degree of accuracy in characterizing the branching of a molecule, making it one of the best indices for measuring branching. The Zagreb index increases proportionally with the degree of branching of the chemical compound. It is also used in selecting chemical compounds with suitable structures and in pharmaceutical modeling [4]. Despite the methodological disparity between the two aforementioned indices, they complement each other in elucidating the molecular concepts of compounds; the strength of this relationship lies in the seamless transition from local (topical) interpretation to structural organization [5,6].

In this paper, we study some selected alkanes. Each alkane is graphically represented, and a new topology is generated based on the alkane's diagram. We then investigate the correlation between the Wiener and Zagreb I indices and some contrasting molecular properties, in addition to calculating the Pearson correlation coefficient between each indices and the number of induced topological configurations. Through statistical analysis, the results obtained for each indices are compared across some structural features on the generated topological complexity. These results contribute to bridging the gap between topological concepts through diagrams and chemical analysis.

2 Preliminaries

In this section we cover some important definitions and essential basics related to finite group theory, graph theory and topological space.

Definition 1. Let $\Gamma = (V, E)$ be a finite simple graph where V denotes the vertex set and $E \subseteq \{\{u, v\}: u, v \in V, u \neq v\}$ represents the edge set.
A graph G is said to be connected if there exists at least one path between every pair of vertices in V.
A chemical graph is a graph that represents the structure of a chemical compound, where vertices correspond to atoms and edges correspond to chemical bonds between them. Typically, hydrogen atoms are omitted, and the graph is considered simple and connected [7].

Definition 2. For a connected graph Γ, the distance $d(u, v)$ between two vertices $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined as the length of a shortest path joining them. This metric captures the minimal number of edges required to move from one vertex to another within the graph structure [12].

Within this paper, the notion of distance serves as a global structural indicator.

Definition 3. Let $\Gamma = (V, E)$ be a connected graph of order $n = |V|$, and let $u, v \in V$ with $u \neq v$. The Wiener index of G is defined as the sum of the shortest-path distances between all unordered pairs of distinct vertices in G. It is expressed as

$$W(\Gamma) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{i} d(v_i, v_j),$$

With $d(v_i, v_j)$ defined as the least number of edges connecting $v_i \; and \; v_j$ [11].

Definition 4. Let Γ be a connected graph with vertex set $V(\Gamma)$. The Wiener index of Γ, denoted by $W(\Gamma)$, is defined as

$$W(\Gamma) = \sum_{\{u,v\} \subseteq V(\Gamma)} d(u, v)$$

The aggregation is performed over all non-repeating pairs of separate vertices.
This index measures the aggregate distance distribution of the graph and reflects its global spread. In the present study, $W(\Gamma)$ is computed for graphs constructed from specific group relations in order to examine how structural dispersion influences the number of induced topologies [7].

Definition 5. For a connected graph Γ , the first Zagreb index is defined as the summation of the squared degrees of all its vertices. It can be mathematically expressed as follows:

$$M_1(\Gamma) = \sum_{u \in V(\Gamma)} \deg(u)^2$$

In this context, $\deg(u)$ signifies the degree of vertex u, which corresponds to the total number of edges incident to that vertex [10].

Definition 6. Pearson's correlation coefficient is a parametric statistical measure used to estimate the strength and direction of the linear relationship between two continuous quantitative variables (X and Y). This coefficient reflects the effect size of the relationship strength, and its application requires the data to follow a normal distribution. It is mathematically expressed as follows [8]:

$$\mathfrak{r}_p = \frac{n\sum(XY) - \sum(X)\sum(Y)}{\sqrt{(n\sum(X^2) - (\sum X)^2)(n\sum(Y^2) - (\sum Y)^2)}}$$

Definition 7. Simple linear regression is a statistical method used to estimate the coefficients $\alpha \; and \; \beta$ of a linear model that describes the relationship between a single dependent variable Y and a single independent variable X, expressed as :

$$Y = \alpha + \beta X,$$

Here, α represents the intercept of the line (the expected value of Y when $X = 0$), and β represents the slope (the expected change in Y for a one-unit change in X[9].

The values α and β are calculated from the following two relationships:

Where,

$$\beta = \frac{n\sum(XY) - (\sum X)(\sum Y)}{n\sum(X^2) - (\sum X)^2} \quad , \quad \alpha = \frac{\sum(Y) - b\sum(X)}{n}$$

3 Molecular graph

Graph theory constitutes the mathematical discipline concerned with the study of discrete structures composed of nodes (vertices) and edges. A graph $\Gamma = (V, E)$ consists of a finite non-empty set V of vertices and a set $E \subseteq \{\{u, v\}: u, v \in V, u \neq v\}$ of edges, where each edge connects a pair of vertices. Two vertices are said to be adjacent whenever an edge joins them.

Graphs are commonly represented diagrammatically: vertices are depicted as points, and edges as line segments joining adjacent vertices. This visual representation provides an intuitive interpretation of structural connectivity without altering the underlying combinatorial structure.

In Figure 1, three graphs $\Gamma_1, \Gamma_2, \Gamma_3$ are shown. While Γ_1 *and* Γ_2 consist of twelve vertices each, Γ_3 contains eleven vertices. The first two graphs represent valid molecular graphs, whereas the third does not satisfy the structural conditions required for a molecular interpretation.

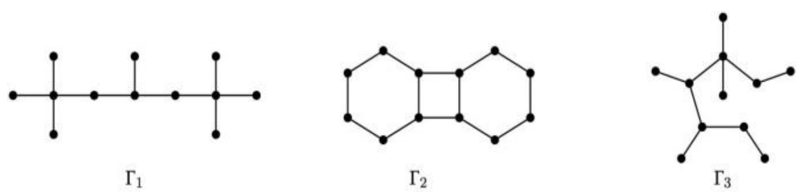

Figure 1. Sample graphs showing that Γ_1 and Γ_2 are molecular graphs, whereas Γ_3 is not.

Graph-theoretical methods have found applications across numerous scientific domains, including computer science, engineering, biology, and especially chemistry. In chemical graph theory, molecular structures are modeled through graphs in which vertices correspond to atoms and edges correspond to chemical bonds. This structural analogy has been extensively documented in the literature [13–20] and further analyzed in review works [21–25].

A molecular graph is necessarily connected and encodes the carbon-atom skeleton of the compound under consideration. In practice, hydrogen atoms and bond multiplicities are often omitted for simplification when focusing on structural topology [16,18].

There also exist graphs that do not correspond to any molecular structure; for instance, Γ_3 in Figure 1, fails to represent a chemically admissible configuration.

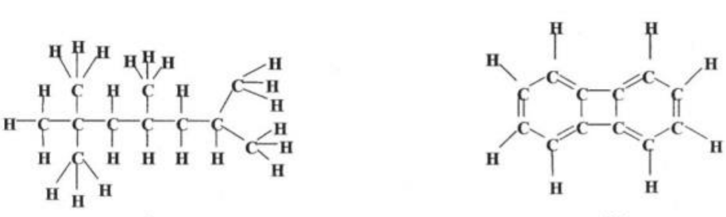

Figure 2. Chemical structures of 2,2,4,6-tetramethylheptane and diphenylene and the associated molecular graphs derived from them.

Example 1: Consider graph Γ_4 shown in Figure 3, which represents the graphical model of the 1,1-dimethylcyclopentane molecule a succession.
A sequence such as 2,3,7,6,5 defines a path of length four connecting vertices 2 and 5. However, this is not the shortest connection between these vertices. The sequence 2,3,4,5 provides a shorter path of length three; therefore, the metric distance satisfies $d_{25} = 3$
Using similar reasoning, all pairwise distances in Γ_4 can be computed. The Wiener index is then obtained by summing distances over all unordered vertex pairs:

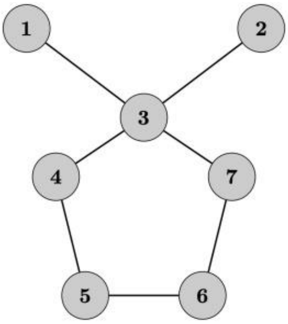

Figure 3. The graph Γ_4 of 1,1-dimethylcyclopemtane.

At this stage, one can validate Γ_4 the results pertaining to graph.

The Distance matrix of the graph:

$$D(\Gamma_4) = \begin{pmatrix} 0 & 2 & 1 & 2 & 3 & 3 & 2 \\ 2 & 0 & 1 & 2 & 3 & 3 & 2 \\ 1 & 1 & 0 & 1 & 2 & 2 & 1 \\ 2 & 2 & 1 & 0 & 1 & 2 & 2 \\ 3 & 3 & 2 & 1 & 0 & 1 & 2 \\ 3 & 3 & 2 & 2 & 1 & 0 & 1 \\ 2 & 2 & 1 & 2 & 2 & 1 & 0 \end{pmatrix}$$

The W-index represents the total summation of distances between every pair of nodes (vertices) in the graph:

$$W(\Gamma) = \sum_{x<y} d(x, y)$$

Substituting the calculated distances yields:

$$W(\Gamma_4) = 39$$

And the Zagreb index

$$M_1(\Gamma_4) = \sum_{u \in V(\Gamma)} \deg(u)^2$$
$$M_1(\Gamma_4) = 34$$

For small graphs, direct summation is feasible. However, for larger molecular structures this method becomes computationally demanding. Consequently, several efficient techniques and algorithmic approaches have been developed to evaluate the Wiener index more effectively [26].

4 Topological Indices and Induced Topology of Selected Alkanes

In this section, the proposed framework is applied to selected alkane molecular graphs in order to investigate the relationship between structural invariants and induced topological complexity.
Each molecule is represented as a simple connected graph in which vertices correspond to carbon atoms and edges correspond to covalent bonds.
For every case study, we:

1. Compute the Wiener indice $W(G)$
2. Compute the first Zagreb indice $M_1(G)$
3. Construct the induced topological space T_G from vertex neighborhoods.

The molecules are arranged in increasing structural complexity:

- $Methane \quad CH_4$
- $Ethane \quad C_2H_6$
- $Propane \quad C_3H_8$
- $Butane \quad C_4H_{10}$
- $Pentane \quad C_5H_{12}$
- $Hexane \quad C_6H_{14}$
- $Heptane \quad C_7H_{16}$
- $Octane \quad C_8H_{18}$
- $Nonane \quad C_9H_{20}$
- $Decane \quad C_{10}H_{22}$

4.1 Computation of Topological Indices

This section is dedicated to calculating the topological indices of the alkanes under study, as well as the topologies induced from the graphs.

Example 2: Topological Analysis of Propane (C_3H_8)

Chemical structure of the molecule

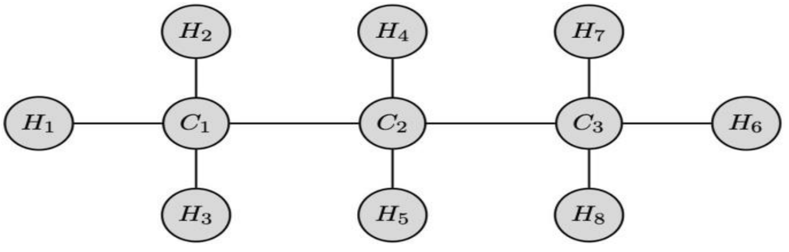

Figure 4. Chemical structure of propane

Conversion to the Molecular Graph: To perform the topological analysis, the chemical structure is transformed into its corresponding molecular graph is constructed by treating carbon atoms as nodes and the C–C bonds as connecting edges, while hydrogen atoms are omitted for simplicity.

Figure 5. Graph representation of the carbon skeleton of propane.

The vertex set and edge set :

$$V = \{v_1\ , v_2\ , v_3\}\ and\ E = \{v_1v_2\ , v_2v_3\}$$

Table 1. The matrix of distances derived from the graphical model of propane

	v_1	v_2	v_3	Sum
v_1	0	1	2	3
v_2	1	0	1	2
v_3	2	1	0	3
				8

The weiner and Zagreb index :

$$W(\Gamma) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{i} d_{ij}$$

$$W(\Gamma) = \frac{1}{2}(8) = 4$$

And

$$M_1(\Gamma) = \sum_{u \in V(G)} \deg(u)^2$$

$$M_1(\Gamma) = 1^2 + 2^2 + 1^2 = 6$$

Compute neighborhoods:
$N(v_1) = \{v_1, v_2\ \}$
$N(v_3) = \{\ v_2, v_3\}$
$N(v_2) = \{v_1, v_2\ , v_3\}$

The Subbasis :

$$S_\Gamma(v) = \{S(v) |\ v \in V(\Gamma)\}$$

$$S(v) = \{v\} \cup N(v)$$

$$S_\Gamma = \{\{v_1\ , v_2\}, \{v_2\ , v_3\}, \{v_1, v_2\ , v_3\}\}$$

$$\beta_\Gamma = \{\cap\, S_i\}$$

$$\beta_\Gamma = \{\{v_1\ , v_2\}, \{v_2\ , v_3\}, \{v_2\}, \{v_1, v_2\ , v_3\}\}$$

Hence :

$$\tau_\Gamma = \{\cup\, \beta_i\}$$

$$\tau_\Gamma = \{\phi, V\ , \{v_1\ , v_2\}, \{v_2\ , v_3\}, \{v_2\}\}$$

Number of open set $|\tau_\Gamma| = 5$
As summarized in Table 2, the values of Wiener and First Zagreb indices vary among the selected alkanes, reflecting their different topological structures.

Table 2. Summary of topological indices and topology sizes for the selected alkanes

Chemical name	Wiener index W	Zagreb index M_1	Size of topology, $\lvert\tau_\Gamma\rvert$
Methane	0	0	2
Ethane	1	2	3
Propane	4	6	5
Butane	10	10	8
Pentane	20	14	13
Hexane	35	18	21
Heptane	56	22	34
Octane	84	26	55
Nonane	120	30	89
Decane	165	34	144

Structural Observation

A comparative examination of the calculated topological indices reveals that structural branching produces a clear effect on the descriptors.
This behavior confirms that the Wiener index functions as a global, distance-based measure, while the Zagreb I index functions as a degree-based measure of local branching density and connectivity concentration.

Topological Interpretation:
The number of open sets varies according to molecular geometry.
Linear structures produce different induced topologies compared to branched structures, even when the number of vertices is identical.
Thus, the induced topology reflects structural connectivity rather than merely vertex count.

5 Statistical Relationship Between Structural Indices and Topological Complexity

To investigate the straight-line association linking two factors, one independent the wiener index (X), and the other a dependent variable, the topological basis (Y) of chemical compounds by simple linear regression.

Let

$$X = \{0,1,4,10,20,35,56,84,120,165\}\ Wiener\ index\ values$$
$$Y = \{2,3,5,8,13,21,34,55,89,144\}\ Number\ of\ open\ sets.$$

Table 3. Data for Pearson correlation between the Wiener index and topology size of selected

X	Y	XY	X^2	Y^2
0	2	0	0	4
1	3	3	1	9
4	5	20	16	25
10	8	80	100	64
20	13	260	400	169
35	21	735	1225	441
56	34	1904	3136	1156
84	55	4620	7056	3025
120	89	10680	14400	7921
165	144	23760	27225	20736
$\sum X = 495$	$\sum Y = 374$	$\sum XY = 42062$	$\sum X^2 = 53559$	$\sum Y^2 = 33550$

Pearson correlation $r_w = 0.987733$ Very strong positive correlation.

It is noted that the linear correlation relationship and the Wiener Index are the topological basis for graphical chemical compounds. After representing the ordered pairs in the plane, we obtain the form of the spread (Very strong expulsion relationship). This means that the relationship between both variables is proportional and linear. We estimate the regression line $Y\ on\ X$ by the relationship. The regression model

$$\bar{Y} = 0.8105X - 2.7175$$

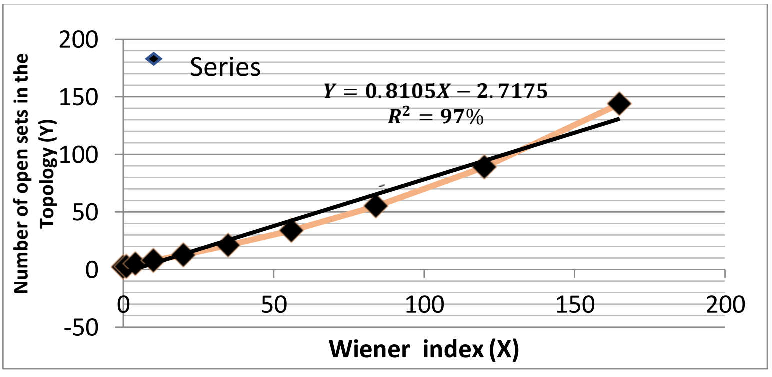

Figure 6. Linear regression between Wiener index values and the number of open sets in the topology

Zagreb index:
While study the linear relationship between two variables, one independent the Zagreb index (X), and the other a dependent variable, the topological basis (Y) of chemical compounds by simple linear regression.
Let

$$X = \{0,2,6,10,14,18,22,26,30,34\} \text{ Zagreb } index\ values$$
$$Y = \{2,3,5,8,13,21,34,55,89,144\}\ Number\ of\ open\ sets$$

Table 4. Data for Pearson correlation between the Zagreb index and topology size of selected

X	Y	XY	X^2	Y^2
0	2	0	0	4
2	3	6	4	9
6	5	30	36	25
10	8	80	100	64
14	13	182	196	169
18	21	378	324	441
22	34	748	484	1156
26	55	1430	676	3025
30	89	2670	900	7921
34	144	4896	1156	20736
$\sum X = 162$	$\sum Y = 374$	$\sum XY = 10420$	$\sum X^2 = 3876$	$\sum Y^2 = 33550$

Pearson correlation: $r_{M_1} = 0.881378$ Strong correlation.
It is noted that the linear correlation relationship and the Zagreb Index are the topological basis for graphical chemical compounds. After representing the ordered pairs in the plane, we obtain the form of the spread (moderate expulsion relationship). This means the relationship between both variables is proportional and linear. We estimate the regression line $Y\ on\ X$ by the relationship.

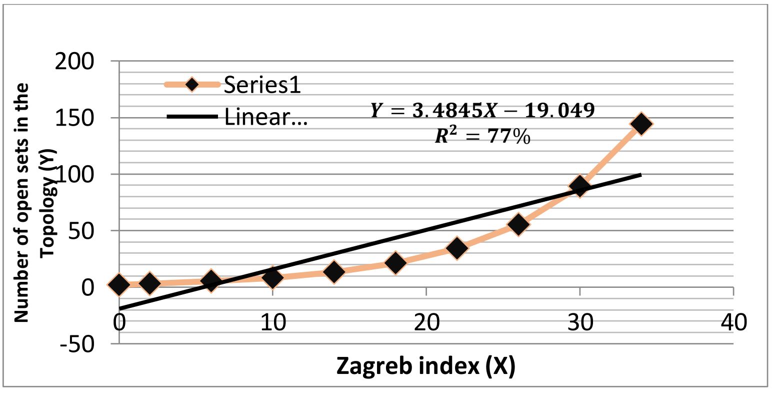

Figure 7. Linear regression between Zagreb index values and the number of open sets in the topology

A deep analytical comparison between W and M_1 based on the diagrams

Table 5. Pearson correlation coefficients between topological indices and the number of open sets

Index	Correlation coefficient	Strength of relationship
Wiener (W)	0.98	Very strong
Zagreb (M)	0.88	strong

The Wiener index describes topology more accurately because it reflects overall changes in molecular shape, while the Zagreb index reflects only localized changes.

Therefore, it can be concluded that the Wiener index provides a more accurate description of the overall topological properties of molecules compared to the Zagreb index, a finding supported by the higher correlation coefficient with the number of open groups in the topology associated with the studied molecules.

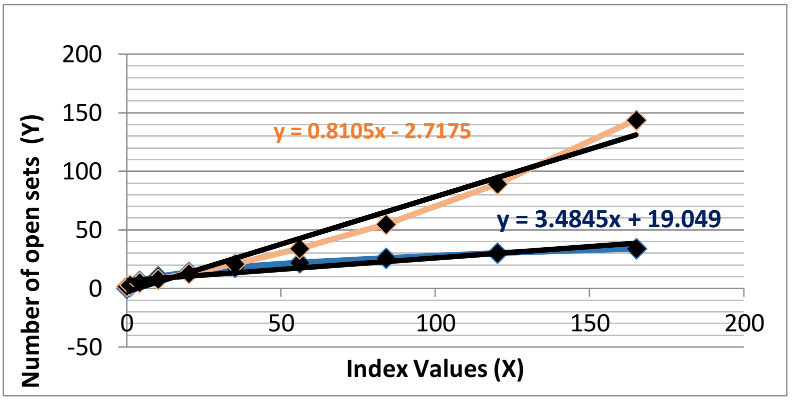

Figure 8. Comparison of linear regression between Wiener and Zagreb indices versus the number of open sets in the topology

6 Conclusion

The results of this study show that the Wiener index has a greater ability to detect and interpret topological changes in organic molecules compared to the Zagreb index. This is due to its reliance on the full-space structure of the chemical diagram, making it sensitive to chain length, the number of rings, and any changes in the overall shape of the molecule.

In contrast, the Zagreb index provides a local representation primarily based on the degrees of atoms. This feature makes the index more suitable for studying local bonds and branching, although it remains less effective in distinguishing molecules with different geometric shapes. Regression plots confirm this observation. It is evident that the correlation coefficient between the number of topological configurations and the numerical values of the Wiener index is higher than that of the Zagreb index for the same topological configurations. Consequently, this indicates that topologies induced from graphs rely more on distance patterns than on vertex degrees. This result provides a complete and comprehensive picture when both indices are employed.

The Wiener index offers a full description of the overall molecular structure, whereas the Zagreb index characterizes only the local structure.
The combined use of these indices strengthens the approach for studying chemical topology and interpreting structural phenomena in compounds.

References

1. A. Bino, M. Murugan, Topological indices of molecular graph and drug design, IJRASET, 2022. https://www.ijraset.com/research-paper/topological-indices-of-molecular-graph-and-drug-design
2. M. H. Akhbari et al., Developing a QSPR model for Alzheimer's drugs using topological indices and M-polynomial, ACS Omega, 16, 1900, 2023.https://doi.org/10.1021/acsomega.3c01900
3. I. Gutman et al., Some recent results in graph theory, 1972. at: https://scholar.google.com/scholar?q=Some+recent+results+in+graph+theory+Gutman+1972
4. S. Wagner, A class of trees and its Wiener index, Acta Appl. Math., 91, 119–132, 2006. https://doi.org/10.1007/s10440-006-9050-4
5. G. Ringle, A. Llado, Another tree conjecture, Bull. Inst. Comb. Appl., 18, 83–85, 1996.https://scholar.google.com/scholar?q=Another+tree+conjecture+Ringle+Llado
6. N. Trinajstić, Chemical graph theory, Wiley, 1986.https://www.wiley.com/en-us/Chemical+Graph+Theory-p-9780849351872
7. M. H. Kutner et al., Applied linear statistical models, Int. Ed. https://www.mheducation.com/highered/product/applied-linear-statistical-models-kutner/M9780073108742.html
8. A. G. Asuero, J. Martín, Pearson's correlation coefficient: A review, J. Appl. Math. Phys., 10, 3245–3258, 2022. https://doi.org/10.4236/jamp.2022.1011218
9. C. Kuratowski, Topology, Warsaw, 1952. https://scholar.google.com/scholar?q=Kuratowski+Topology+1952
10. I. Gutman, N. Trinajstić, Graph theory and molecular orbitals, Chem. Phys. Lett., 17, 535–538, 1972. https://doi.org/10.1016/0009-2614(72)85099-1
11. H. Wiener, Structural determination of paraffin boiling points, J. Am. Chem. Soc., 69, 17–20, 1947.https://doi.org/10.1021/ja01193a005
12. R. Diestel, Graph theory, Springer, 2017. https://link.springer.com/book/10.1007/978-3-662-53622-3
13. A. T. Balaban, Chemical applications of graph theory, Academic Press, 1976.https://www.sciencedirect.com/book/9780120774504/chemical-applications-of-graph-theory
14. A. Graovac et al., Topological approach to the chemistry of conjugated molecules, Springer, 1977. https://link.springer.com/book/10.1007/978-3-642-61969-4
15. D. Bonchev, Information theoretic indices for characterization of chemical structure, 1983.https://scholar.google.com/scholar?q=Bonchev+Information+Theoretic+Indices+1983
16. I. Gutman, O. E. Polansky, Mathematical concepts in organic chemistry, Springer, 1986. https://link.springer.com/book/10.1007/978-3-642-70954-8

17. D. Bonchev, D. H. Rouvray, Chemical graph theory: Introduction and fundamentals, Gordon and Breach, 1991. https://www.taylorfrancis.com/books/mono/10.1201/9780203739698
18. N. Trinajstić, Chemical graph theory (2nd ed.), CRC Press, 1992. https://www.crcpress.com/Chemical-Graph-Theory/Trinajstic/p/book/9780849342559
19. J. R. Dias, Molecular orbital calculations using chemical graph theory, Springer, 1993. https://link.springer.com/book/10.1007/978-3-642-78020-2
20. M. V. Diudea, O. Ivanciuc, Molecular topology, 1995. https://scholar.google.com/scholar?q=Diudea+Molecular+Topology+1995
21. D. H. Rouvray, Chemical applications of graph theory, Science, 255, 36–43, 1986. https://doi.org/10.1126/science.255.5040.36
22. P. G. Seybold et al., Applications of graph theory in chemistry, J. Chem. Educ., 64, 575–581, 1987. https://doi.org/10.1021/ed064p575
23. A. T. Balaban et al., Topological indices in chemistry, J. Chem. Educ., 65, 304–313, 1988. https://doi.org/10.1021/ed065p304
24. P. J. Hansen, P. C. Jurs, Chemical applications of graph theory, J. Chem. Educ., 65, 574–580, 1988. https://doi.org/10.1021/ed065p574
25. Z. Mihalić, N. Trinajstić, Graph theoretical descriptors in chemistry, J. Chem. Educ., 69, 701–712, 1992. https://doi.org/10.1021/ed069p701
26. S. Nikolić et al., The Zagreb indices revisited, Croat. Chem. Acta, 68, 105–129, 1995. https://hrcak.srce.hr/134119
27. J. R. Rumble, CRC handbook of chemistry and physics, CRC Press, 2023. https://www.routledge.com/CRC-Handbook-of-Chemistry-Physics/Rumble/p/book/9781032015021