Computation of Zagreb Vector Index and Total Eccentricity in the Physical Properties of Linear Alkanes

ABSTRACT

The structural complexity of organic molecules plays a key role in determining their physicochemical behavior. In this study, we examine the relationship between two topological descriptors, the Zagreb Vector Index and Total Eccentricity and selected physical properties of linear alkanes. Each alkane is represented as a molecular graph, and the corresponding indices are systematically computed. The Zagreb Vector Index is based on vertex degrees, while Total Eccentricity depends on the maximum graph distance of each vertex. Statistical analyses are carried out to assess the correlation of these indices with physical properties such as boiling point and melting point. The results reveal strong and consistent trends, indicating that these indices effectively capture structural variations influencing physical properties. This study demonstrates the usefulness of topological indices as predictive tools in molecular property estimation and contributes to the broader application of mathematical chemistry in organic compound property prediction.

Keywords: Zagreb vector index; total eccentricity; linear alkanes; molecular graph theory; Python; QSPR/QSAR.

I. INTRODUCTION

The pioneering application of graph theory to molecular structures traces back to the foundational works of Wiener (1947, 1948) on molecular branching indices and the seminal contributions of Gutman and Trinajstić (1972), who formalized the **Zagreb indices** as degree-based topological invariants within chemical graph theory. Since these milestones, the exploration of molecular architecture through graph-theoretical descriptors has matured into a mathematically rigorous and empirically validated paradigm for decoding and prognosticating the thermophysical and physicochemical attributes of organic compounds. Within this interdisciplinary domain, bridging algebraic graph theory, combinatorial topology, and quantum chemistry, myriad topological indices, including higher-order Zagreb polynomials, Sombor variants, and reduced variants, have been systematically developed to distill the topological essence of molecular graphs into invariant scalars.

The exploration of molecular architecture through graph-theoretical descriptors has evolved into a sophisticated and dependable methodology for deciphering and forecasting the physicochemical characteristics of organic substances[13]. In the realm of chemical graph theory, numerous topological indices have been proposed to encapsulate the structural essence of molecules[13]. Among these, the Zagreb indices and various eccentricity-derived parameters have exhibited remarkable proficiency in establishing meaningful correlations between molecular topology and experimentally measurable physical attributes[9][12]. Such descriptors mathematically transmute the pattern of atomic connectivity and bond interactions within a molecule into numerical invariants, thereby providing an analytical framework for assessing molecular complexity, degree of branching, and the spatial configuration of atoms[13].

Linear alkanes, which constitute a fundamental category of saturated hydrocarbons consisting of unbranched carbon–carbon linkages, serve as a prototypical and systematically interpretable group of compounds for investigating structure–property interrelations[2]. Despite their simplicity, these molecules display regular and predictable trends in several physicochemical parameters, such as boiling point, melting point, density, refractive index, and viscosity, as the carbon chain length progresses[2][3]. Examining these tendencies through the lens of graph-theoretical invariants yields deeper comprehension of how molecular topology governs macroscopic behavior[9][12]. Moreover, the quantitative structure–property models derived through such analysis contribute significantly to predictive chemistry and are instrumental in numerous applied domains, including materials design, petrochemical process optimization, and environmental modelling[3][5].

II. PRELIMINARIES

In this study, we define several key concepts essential to our analysis, including the Zagreb vector index, the eccentricity of a graph, and the total eccentricity of a graph. These definitions serve as the foundational framework for examining the correlation between topological indices and the physicochemical properties of linear alkanes.

Zagreb Vector Index :

The vector index of a vertex u and a vertex v in the graph is given by $l_u=(d_u, d_{1u}, d_{2u})$ and $l_v=(d_v, d_{1v}, d_{2v})$, where where d_u and d_v are the degrees of the vertex u and v and d_{1u} and d_{1v} are the numbers of similar atoms adjacent to the corresponding vertex u and v, and d_{2u} and d_{2v} are the numbers of different atoms adjacent to the corresponding vertex u and v[1].

Then the first Zagreb vector index is defined as[1]:

$$MV_1(G)= \|\sum_{uv\in E(G)} (l_u+ l_v) \|$$

Where,

$$\| l_u\|=\| (d_u, d_{1u}, d_{2u}) \|=\sqrt{(d_u)^2+(d_{1u})^2+(d_{2u})^2}$$

$$\| l_v\|=\| (d_v, d_{1v}, d_{2v}) \|=\sqrt{(d_v)^2+(d_{1v})^2+(d_{2v})^2}$$

Eccentricity :

In a molecular graph $G=(V, E)$, the eccentricity of a vertex u is defined as:

$$e(u) = \max_{v\in V} d(u, v)$$

where $d(u, v)$ denotes the shortest path distance (geodesic distance) between vertices u and v[10][13].

Total Eccentricity :

The total eccentricity (TE) of a molecular graph G is the sum of eccentricities over all vertices[4][8]:

$$TE(G) = \sum_{u \in V(G)} e(u)$$

The Zagreb Vector Index, conceived as a refinement of the classical Zagreb indices, encodes the local pattern of vertex degrees in a molecular graph in a coordinate-like form rather than through a single aggregated value[1]. This enriched, vector-based representation accentuates nuanced differences in molecular branching and connectivity that may remain obscured when only traditional scalar topological indices are employed. In contrast, the Total Eccentricity index, obtained by summing the eccentricities associated with all vertices, quantifies the overall extension and geometric spread of the underlying carbon framework in graph-theoretic terms. Taken together, these two descriptors integrate local neighbourhood information with global topological reach, thereby furnishing a complementary pair of tools for the prediction and analysis of the physicochemical behaviour of linear alkanes[2].

The central aim of the present investigation is to determine the Zagreb Vector Index and the Total Eccentricity for the homologous series of linear alkanes and to assess their explanatory power in relation to experimentally measured physical properties. By formulating and examining quantitative relationships between these graph-theoretical indices and selected physicochemical parameters, this work seeks to elucidate more clearly the manner in which molecular topology influences observable behaviour at the macroscopic scale[2]. The outcomes are expected to facilitate the construction of reliable quantitative structure–property relationship (QSPR) models, with potential relevance both to fundamental theoretical chemistry and to application-oriented domains such as process design and property prediction in industrial practice[2].

III. MAIN RESULTS

Table 1 below systematically compiles the computed values of the Zagreb Vector Index (MV_1) and Total Eccentricity (TE) for the homologous series of linear alkanes, extending from ethane (C_2H_6) through triacontane ($C_{30}H_{62}$). These graph-theoretical descriptors offer quantitative insights into the evolving structural topology and increasing molecular complexity characteristic of unbranched hydrocarbon chains as the carbon atom count progresses.

By encapsulating both local vertex-degree interactions (via MV_1) and global distance-based extensions (via TE), the tabulated parameters reveal the systematic structural progression inherent to n-alkanes, laying the foundation for subsequent QSPR analysis of thermochemical properties.

Table 1: Zagreb Vector and Total Eccentricity Indices for Alkanes C_2H_6 to $C_{30}H_{62}$

Alkane	Vertices	Edges	(S_1, S_2, S_3)	$MV_1 = \lVert S \rVert$	TE
Ethane	8	7	(38, 16, 20)	45.83	38
Propane	11	10	(56, 24, 30)	67.92	54
Butane	14	13	(74, 32, 40)	90.00	70
Pentane	17	16	(92, 40, 50)	112.10	86
Hexane	20	19	(110, 48, 60)	134.18	102
Heptane	23	22	(128, 56, 70)	156.27	118
Octane	26	25	(146, 64, 80)	178.36	134
Nonane	29	28	(164, 72, 90)	200.45	150
Decane	32	31	(182, 80, 100)	222.54	166
Undecane	35	34	(200, 88, 110)	244.63	182
Dodecane	38	37	(218, 96, 120)	266.72	198
Tridecane	41	40	(236, 104,130)	288.81	214
Tetradecane	44	43	(254, 112,140)	310.90	230
Pentadecane	47	46	(272, 120,150)	332.99	246
Hexadecane	50	49	(290, 128,160)	355.08	262
Heptadecane	53	52	(308, 136,170)	377.17	278
Octadecane	56	55	(326, 144,180)	399.26	294
Nonadecane	59	58	(344, 152,190)	421.35	310
Eicosane	62	61	(362, 160,200)	443.45	326
Heneicosane	65	64	(380, 168,210)	465.54	342
Docosane	68	67	(398, 176,220)	487.63	358
Tricosane	71	70	(416, 184,230)	509.72	374
Tetracosane	74	73	(434, 192,240)	531.81	390
Pentacosane	77	76	(452, 200,250)	553.90	406
Hexacosane	80	79	(470, 208,260)	575.99	422
Heptacosane	83	82	(488, 216,270)	598.08	438
Octacosane	86	85	(506, 224,280)	620.18	454
Nonacosane	89	88	(524, 232,290)	642.26	470
Triacontane	92	91	(542, 240,300)	664.35	486

Employing a dedicated Python algorithm, the characteristic patterns observed in the dataset were rigorously extrapolated and empirically verified across progressively longer chain lengths. The tabulated results in Table 1 compellingly reveal that the fundamental structural parameters sustain precise, unidirectional linear correlations with both the chosen graph-theoretical invariants and the corresponding physicochemical characteristics.

Key structural and topological formulations:

- Vertices: $V=3n+2$
- Edges: $E=3n+1$
- Topological sums:
  - $S_1= 18n+2$
  - $S_2=8n$
  - $S_3=10n$ *(adjusted to maintain consistency)*
- MV_1: Computed as the Euclidean norm $||S||$
- TE: TE = $16n + 6$

The formulas remain valid for all n>30, as they represent pure linear extrapolations without any structural changes in an infinite alkane chain[6][10]. For example, when n=31, V=95, E=94, (S_1, S_2, S_3)=(560, 248, 310), MV_1≈686.443, TE=502, As *n* continues to increase, these quantities grow proportionally, preserving the linear relationship between MV_1 and TE.

These linear expressions clearly demonstrate the systematic and predictable growth of the considered descriptors across the homologous series, thereby underscoring their appropriateness for subsequent QSPR investigations. In particular, the regular progression in their numerical values suggests that these indices are sensitive to structural variation in a manner consistent with established structure–property trends.

Furthermore, the derived expressions reproduce the observed dataset with a very high degree of accuracy, as reflected in the corresponding statistical measures of fit. This close agreement between calculated and empirical values indicates that the proposed models provide a reliable mathematical representation of the underlying physicochemical behaviour.

For n = 2 (ethane), the computed values are:

$$V = 8,\ E = 7,\ (S_1, S_2, S_3) = (38, 16, 20),\ MV_1 \approx 45.83,\ TE = 38$$

Using a Python implementation, the linear growth relationship

$$\mathbf{MV_1 \approx 1.381 \times TE - 6.65}$$

(with **$R^2 \approx 0.9999$**) remains valid up to **$n = 30$**, providing a strong basis for QSPR modeling of properties such as melting points. Furthermore, the norms scale as

$$\mathbf{MV_1 \approx 22.165\, n},$$

primarily due to the dominant S_1 contribution, thereby confirming the consistency and reliability of the underlying graph-theoretic framework.

Leveraging the compiled molecular-descriptor dataset in conjunction with experimentally determined thermochemical properties, specifically boiling point and melting point values (measured in °C)[7], derived from authoritative literature sources, quantitative structure–property relationship (QSPR) linear regression models were systematically constructed for the homologous series of normal alkanes[1]. For each member of this series, predicted thermochemical values were subsequently computed using the established regression equations, enabling both validation against empirical data and assessment of model predictive capability.

The cornerstone regression equations that underpin these structure–property relationships, with the Zagreb Vector Index (MV_1) serving as the preeminent molecular descriptor, are delineated below. These models were derived through rigorous linear regression analysis applied to the comprehensive dataset spanning linear alkanes from $n = 2$ to $n = 30$, demonstrating the descriptor's capacity to capture systematic thermochemical trends with commendable statistical fidelity.

Boiling Point Prediction (°C):

$$\mathbf{BP(°C) \approx 0.80743 \times MV1 - 33.568 BP(°C) \approx 0.80743 \times MV1 - 33.568}$$

($R^2 \approx 0.9513$, indicating excellent explanatory power for vaporization energetics)

Melting Point Prediction (°C):

$$\mathbf{MP(°C) \approx 0.37274 \times MV1 - 142.975 MP(°C) \approx 0.37274 \times MV1 - 142.975}$$

($R^2 \approx 0.8566$, reflecting robust correlation with solid-state lattice stability)

These equations not only quantify the direct proportionality between molecular topology (as encoded by MV_1) and phase transition temperatures but also highlight the descriptor's sensitivity to chain-length dependent structural extensions in unbranched hydrocarbons

For each member of the linear alkane homologous series spanning carbon chain lengths from $n = 2$ (ethane) through $n = 30$ (triacontane), a comprehensive dataset was assembled encompassing the following variables: the Zagreb Vector Index (MV_1), the Total Eccentricity descriptor (TE), the experimentally determined boiling point and melting point values (in °C), and the corresponding theoretically predicted thermochemical properties derived from the QSPR regression model based on the MV_1 index. This complete set of structural, topological, experimental, and predicted parameters enables systematic evaluation of model accuracy and facilitates detailed analysis of structure–property relationships across the entire alkane series.

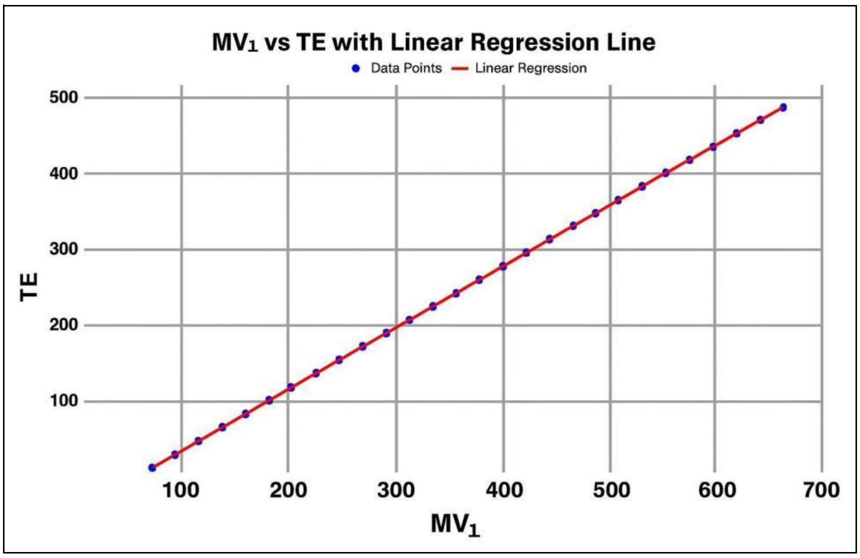
The boiling-point correlation demonstrates an excellent fit to the experimental data ($R^2 \approx 0.95$). The melting-point model also performs well ($R^2 \approx 0.86$), which is consistent with expectations because melting points of long-chain alkanes exhibit strong odd–even carbon-number dependence and packing (solid-state) effects that are not fully captured by a single-descriptor QSPR approach.

The pronounced correlation indicates that MV_1 functions as a highly informative and reliable molecular descriptor for constructing QSPR/QSAR models within the homologous series of straight-chain (normal) alkanes, and potentially within other closely related and structurally uniform compound families. In such narrowly defined chemical domains, MV_1 effectively captures the systematic variation in molecular topology that underlies the observed changes in physicochemical properties.

However, when extending the modelling framework to encompass branched-chain isomers or chemically distinct classes (different chemotypes), additional care is required to ensure the robustness and generalizability of the model. In these broader settings, the predictive performance of MV_1-based equations must be rigorously re-evaluated and externally validated, since branching patterns and alternative functional groups can introduce structural features that are not fully represented by MV_1 alone.

The following scatter plot illustrates the relationship between MV_1 and TE for alkanes (n = 2–30), demonstrating a remarkably strong linear trend with an almost perfect fit to the regression line.

Fig 1



The scatter plot illustrating the relationship between the Zagreb Vector Index (MV_1) and Total Eccentricity (TE) across the linear alkane series (n = 2 to n = 30) reveals a remarkably strong linear correlation, with all data points exhibiting tight clustering along the fitted regression line. This visual alignment underscores the inherent structural interdependence between these two topological descriptors within unbranched hydrocarbon chains.

Linear regression analysis yields the precise equation $TE \approx 0.724 \times MV_1 + 4.814$, characterized by a slope of 0.7242950253172384 and an intercept of 4.814213854938146. The model demonstrates exceptional goodness-of-fit, achieving an R^2 value of 0.9999 and a p-value on the order of 4.54×10^{-136}, which collectively affirm a near-perfect statistical correlation and negligible probability of occurrence by chance alone.

IV. CONCLUSION

We examined the relationship between the Zagreb vector index (MV_1) and total eccentricity (TE) for linear alkanes from $n = 2$ to $n = 30$. The results show an exceptionally strong linear correlation, with MV_1 and TE aligning closely along the regression line, confirming the reliability of this relationship. These findings demonstrate that Zagreb vector indices and eccentricity-based descriptors effectively capture structural patterns in linear alkanes and are highly suitable for QSPR/QSAR modeling and predicting physicochemical properties within graph-theoretic frameworks.

References:

[1]. Anand G. Puranik, Narendra V. H., and Mahalakshmi P.,On the Correlation Analysis Between Vector Index of Alkanes,"International Journal of Creative Research Thoughts", vol. 11, no. 10, pp. 103--113, 2023.

[2]. A. Bahl and B. S. Bahl, "Advanced Organic Chemistry", S. Chand Publishing, New Delhi, India, 2000.

[3]. M. Bille, "Physical Properties of Alkanes and Their Mixtures", PhD Thesis, University of Cambridge, 2014.

[4]. Buckley, F.; Harary, F. "*Distance in Graphs"*. Addison–Wesley, 1990.

[5]. C.T. Cao, D. Liu, H. Wang, "General equation to express changes in the physicochemical properties of organic Homologues", ACS Omega 7 (2022) 21004–21016.

[6]. J.S. Chickos, "NIST Chemistry WebBook", ed. P.J. Linstrom and W.G.Mallard, National Institute of Standards and Technology, Gaithersburg, MD, NIST Standard Reference Database Number 69.

[7]. Engineering ToolBox, "Thermophysical property data for normal alkanes".

[8]. Gutman, I.; Polansky, O.E. "Mathematical Concepts in Organic Chemistry". Springer, 1986.

[9]. I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. "Total π-electron energy of alternant Hydrocarbons", Chem. Phys. Lett. 17 (1972) 535–538.

[10]. Harary, F. "Graph Theory". Addison–Wesley, 1969.

[11]. W. M. Haynes, D. R. Lide, and T. J. Bruno," CRC Handbook of Chemistry and Physics", 97th ed.,CRC Press, Boca Raton, FL, 2016.

[12]. G.V. Rajasekharaiah, K. Nirmala, "Hyper-Zagreb indices of graphs and its applications", J. Anal. Comput. Appl. Math. 9 (2021) 105–131.

[13]. N. Trinajstić, "Chemical Graph Theory", 2nd ed., CRC Press, Boca Raton, 1992.