Biointerface Research in Applied Chemistry
Open-Access Journal (ISSN: 2069-5837)

Article
Volume 16, Issue 4, 2026, 115
https://doi.org/10.33263/BRIAC164.115

Quantitative Analysis of Status-Based Topological Indices for Predictive Modeling of Molecular Properties

Jalappa Meti [1], Boregowda H. S [1,*], Vijendra N [1]

[1] Department of Studies and Research in Mathematics, Tumkur University, Jnanasiri Campus, Bidarakatte, Tumkuru-572118, Karnataka, India; metijalappa@g-mail.com (J.M.); bgsamarasa@g-mail.com (B.H.S.); vijendran397@g-mail.com (V.N.);
* Correspondence: bgsamarasa@g-mail.com;

Received: 5.02.2026; Accepted: 27.05.2026; Published: 15.08.2026

Abstract: Polycyclic aromatic hydrocarbons (PAHs) are a class of aromatic molecules consisting of multiple fused benzene rings. In chemical graph theory, topological indices play a crucial role in exploring structure-property relationships and in forecasting physicochemical behavior and biological responses of PAH compounds. Such indices are extensively utilized in computational chemistry, drug discovery, and quantitative structure-property relationship (QSPR) modeling. In this work, we investigate the predictive capability of several status-based topological indices, namely the First Status Connectivity index $S_1(G)$, Second Status Connectivity index $S_2(G)$, Nirmala Status index $SN(G)$, Forgotten Status index $SF(G)$, Status Sombor index $SSO(G)$, and Status Elliptic Sombor index $SESO(G)$ for a selected set of 38 high-priority PAHs. Among the studied descriptors, the Nirmala Status index exhibits the highest predictive accuracy, particularly for molar refractivity and polarizability ($R \approx 0.979$). Furthermore, the variation in the best-performing index across linear, quadratic, and cubic regression models is illustrated in comparative plots based on minimum RMSE values, providing greater clarity and interpretation. The results demonstrate that status-based indices effectively capture long-range structural and electronic characteristics of PAHs. This study highlights the potential of distance-based topological descriptors as reliable tools in QSPR modeling.

Keywords: status sombor index; molecular descriptors; regression analysis; MSC**:** 05C09, 05C90, 62J05.

1. Introduction

Chemical Graph Theory (CGT) has emerged as a powerful mathematical discipline that models molecular structures using graph-theoretic concepts, where atoms correspond to vertices and chemical bonds correspond to edges [1, 2]. This abstraction preserves essential molecular features and enables the construction of numerical descriptors known as topological indices (TIs). These indices encode structural information into real numbers and have become indispensable in Quantitative Structure–Property Relationship (QSPR) and Quantitative Structure Activity Relationship (QSAR) studies because of their computational efficiency and predictive reliability [3, 4]. Over the last five decades, topological indices have played an important role in correlating molecular graphs with experimentally measurable physicochemical, thermodynamic, and biological properties.

A substantial portion of classical topological indices is degree-based. These include the Zagreb indices, Randić connectivity index [5,6], the harmonic index, and atom–bond connectivity measures. Recently, the introduction of the Sombor index by Gutman [7, 8] revitalized research on degree-dependent descriptors due to its geometric interpretation in terms of Euclidean distances between vertex degrees. This was followed by new variants, such as the elliptic Sombor index [9], the Euler Sombor index [10], and refined generalizations applied to extremal graph theory and molecular modeling [11]. Numerous studies have confirmed the strong predictive ability of Sombor-type descriptors for physicochemical properties, including boiling point, molar volume, entropy, refractivity, and thermodynamic stability across many chemical families [12, 13].

However, degree-based indices inherently capture only local connectivity information. In contrast, many molecular phenomena, such as aromaticity, resonance stabilization, π-electron delocalization, and long-range electron communication, are strongly influenced by global structural features. These effects are especially prominent in polycyclic aromatic hydrocarbons (PAHs), benzenoid structures, graphene fragments, fullerenes, and π-conjugated organic compounds. For such molecules, global topological distances, long-range cyclic interactions, and molecular compactness play crucial roles in determining physicochemical behavior [14]. As a result, there has been increasing interest in distance-based descriptors that capture global structural characteristics more effectively.

Among distance-based descriptors, the status (or transmission) of a vertex, defined as the sum of distances from a vertex to all other vertices, has significant chemical relevance [15]. Status values reflect vertex centrality, accessibility, electronic communication routes, and participation in resonance structures. Motivated by these properties, several status-based indices have been introduced, including the First Status Connectivity index and the Second Status Connectivity index [16], the Nirmala Status index [17], and the Forgotten Status index [18]. More recent developments include hybrid descriptors combining structure and Euclidean geometry, such as the Status Sombor index [19] and the Status Elliptic Sombor index [20]. These indices integrate long-range structural information with weighted connectivity, potentially outperforming classical degree-based indices.

The chemical motivation for status-based indices is profound. In PAHs and other conjugated systems, electron delocalization depends heavily on long-range connectivity patterns and ring fusion topologies. Aromatic stabilization, Clar sextet distribution, cyclic delocalization, and resonance energy are influenced by the extent to which electrons can move across the molecular framework [21]. Distances in the molecular graph reflect electron mobility and delocalization length, which are vital for understanding aromaticity. For example, peripheral and internal vertices in coronene, ovalene, benzo[a]pyrene, and chrysene have significantly different distance environments despite having identical degrees, a distinction that status-based indices capture precisely [22].

Additionally, electrophilic aromatic substitution (EAS), a key reaction in aromatic chemistry, strongly depends on electron density distribution and global conjugation patterns. These electronic distributions correlate with topological distances rather than local degrees alone [23]. Recent computational chemistry studies reveal that thermodynamic properties such as enthalpy of formation, molar refractivity, heat capacity, and polarizability are governed by global shape and molecular compactness [24]. As PAHs grow larger and more complex, degree-based indices lose discriminative power, while status-based descriptors maintain sensitivity to both branching and extended resonance structures [25, 26].

Moreover, topological resonance energy, an important aromatic stability index, has been shown to depend on long-range graph distances and global connectivity patterns [27]. This reinforces the idea that status-based measures can better capture the subtleties of molecular stability and aromaticity than classical indices.

Previous studies on QSPR/QSAR modeling have primarily focused on degree-based topological indices, such as the Zagreb indices, Randić index, and Sombor-type descriptors, demonstrating their usefulness in predicting the physicochemical properties of PAHs [28-31]. More recent works introduced several status-based indices, including $S_1(G)$, $S_2(G)$, $SN(G)$, $SF(G)$, $SSO(G)$, and $SESO(G)$ [32-36], but these studies have mainly addressed mathematical properties and structural characteristics rather than their predictive chemical performance.

However, the literature lacks a unified and systematic comparison of status-based indices within a QSPR/QSAR regression framework. No existing study simultaneously evaluates all major status descriptors or examines their behavior under linear, quadratic, and cubic regression models for a consistent dataset of PAHs. Moreover, the chemical relevance of status-based indices, especially their ability to capture long-range interactions and electron delocalization, has not been quantitatively validated [37].

The present work fills this gap by providing the first comprehensive predictive assessment of six major status-based indices for 38 high-priority PAHs. Through correlation analysis, regression modeling, RMSE comparison, and graphical evaluation, this study identifies $SN(G)$ as the most effective descriptor. Thus, the work establishes the practical value of status-based indices in QSPR modeling and highlights their predictive advantage over classical degree-based measures.

2. Materials and Methods

Let G be a connected graph comprising n vertices and m edges, with vertex and edge sets denoted by $V(G)$and $E(G)$, respectively. The status (or transmission) of a vertex $u \in V(G)$, in a graph, is defined as the sum of the shortest path distances from that vertex to all other vertices in the graph, denoted by σ(u), and is defined as [23]:

$$\sigma(u) = \sum_{v \in V(G)} d_G(u, v)$$

The considered status-based indices, such as the First status connectivity index $(S_1(G))$and Second Status Connectivity index $(S_2(G))$ [16], Nirmala Status index $(SN(G))$ [17], the Forgotten Status index $(SF(G))$ [34], Status Sombor index $(SSO(G))$ [35] and Status Elliptic Sombor index $(SESO(G))$ [36] are defined as follows:

$$S_1(G) = \sum_{uv \in E(G)} (\sigma(u) + \sigma(v))$$

$$S_2(G) = \sum_{uv \in E(G)} \sigma(u)\sigma(v)$$

$$NS(G) = \sum_{uv \in E(G)} \sqrt{(\sigma(u) + \sigma(v))}$$

$$FS(G) = \sum_{uv \in E(G)} (\sigma(u)^2 + \sigma(v)^2)$$

$$SSO(G) = \sum_{uv \in E(G)} \sqrt{(\sigma(u)^2 + \sigma(v)^2)}$$

$$SESO(G) = \sum_{uv \in E(G)} (\sigma(u) + \sigma(v)) \sqrt{(\sigma(u)^2 + \sigma(v)^2)}$$

The molecular graph structures for the 38 PAHs analyzed in this study were derived using the approach reported by Kiran *et al.* [39]. According to this representation, atoms are mapped to the vertices of the graph, while the bonds between them correspond to edges. Vertex status values were computed in Python using NetworkX by deriving all-pairs shortest-path distances, and the resulting distance matrices were used to calculate the indices $S_1(G)$, $S_2(G)$, $SF(G)$, $SSO(G)$, $SN(G)$, and $SESO(G)$. Physicochemical property data (Table 2) were obtained from [39]. Linear, quadratic, and cubic regression models were fitted using least-squares estimation in Excel. Model performance was evaluated using correlation coefficients, RMSE, and the F-statistic, ensuring statistically sound and reproducible analysis. Standard regression assumptions were verified; the scatter plots showed clear monotonic trends confirming linearity, independence of errors was ensured since each PAH is structurally distinct, residuals exhibited small standard errors, indicating approximate normality, constant variance was supported by the uniform spread of observed-predicted values, and model adequacy was confirmed by high R^2, low RMSE, and statistically significant F-values. Overall, the cubic model provided the best fit, and $SN(G)$ consistently showed superior predictive performance.

High R^2 values obtained from polynomial regression may be artificially inflated due to model complexity; therefore, model validity was assessed using RMSE, adjusted R^2, and cross-validation rather than relying on R^2 alone.

Classical descriptors were not included as benchmarks because the objective of this study was to evaluate the comparative behavior of status-based indices alone, which has not been previously assessed. Earlier studies already provide extensive benchmarking of classical indices; therefore, re-evaluating them was beyond the scope of the present study.

To contextualize the performance of status-based descriptors, $SN(G)$ and related status indices were compared at the correlation level with established distance-based indices such as the Wiener, Szeged, Harary, Balaban J, and Estrada indices.

The dataset comprises 38 PAHs, which is relatively small for higher-order modeling; therefore, cross-validation and adjusted R^2 were used to control overfitting and ensure model reliability.

2.1. Regression modeling with standardized predictors.

To address scale differences and multicollinearity arising from polynomial regression, the Nirmala Status index was standardized prior to model fitting. Let:

$$Z = \left(S_N(G) - \bar{}\left(S_N(G)\right)\right) / s_{(S_N)}$$

where $\bar{}\left(S_N(G)\right)$ and $s_{(S_N)}$ denote the mean and standard deviation of $S_N(G)$, respectively. Linear, quadratic, and cubic regression models were then constructed using Z, Z^2, and Z^3 as predictors. This reparameterization improves numerical stability and substantially reduces variance inflation without altering fitted values, correlation strength, or predictive accuracy.

3. Results and Discussions

The regression analysis is carried out using linear, quadratic, and cubic models of the form:

$$y = a + b_1 x_1 \quad (1)$$
$$y = a + b_1 x_2 + b_2 x_2^2 \quad (2)$$
$$y = a + b_1 x_3 + b_2 x_3^2 + b_3 x_3^3 \quad (3)$$

Where y denotes the response variable, a is the intercept term, and $b_i (i = 1,2,3)$ are the estimated coefficients corresponding to the explanatory variables x_i. The statistical measures n, r, SE, and F respectively denote the number of observations, the linear correlation coefficient, the standard error of the estimate, and Fisher's statistic for testing overall model significance.

The standard error (SE) quantifies the spread of the observed values around the fitted regression line and is computed as:

$$SE = \sqrt{\frac{1}{n-k} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (4)$$

Where k is the number of model parameters (including the intercept), y_i represents the actual observations, and $\hat{y}_i$ is the corresponding model-predicted value. Lower SE values indicate a tighter model fit.

The predictive capability of the regression models is further assessed using the Root Mean Square Error (RMSE), which measures the average magnitude of deviation between actual and estimated outputs:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (5)$$

In this expression, n represents the total sample size, y_i is the observed response for the i^{th} data point, and $\hat{y}_i$ denotes the corresponding predicted value from the model. A smaller RMSE value reflects stronger predictive accuracy.

The standardized formulation confirms that the strong correlations observed between $S_N(G)$ and physicochemical properties are intrinsic and not driven by numerical scaling. The consistently low VIF values validate the robustness of the regression models. All numerical computations and regression analyses were performed using Maxima and Python [29, 34].

The cross-validation results indicate that although cubic models achieve slightly lower training RMSE, the improvement in CV-RMSE is marginal compared to quadratic models, suggesting mild overfitting. Therefore, quadratic models provide a better balance between predictive accuracy and generalizability."

3.1. Regression procedure and model validation.

To improve methodological transparency and ensure reproducibility, the regression analysis was carried out using a structured procedure. For each physicochemical property, three types of regression models—linear, quadratic, and cubic—were developed using the least-squares estimation method. The models are defined as follows: linear model ($Y = a + b_1X$), quadratic model ($Y = a + b_1X + b_2X^2$), and cubic model ($Y = a + b_1X + b_2X^2 + b_3X^3$), where X represents the topological index, and Y denotes the physicochemical property.

Before model fitting, all predictor variables were standardized to have zero mean and unit variance using the transformation $Z = (X - \mu) / \sigma$, where μ and σ are the mean and standard

deviation of X. This step minimizes scaling effects and reduces multicollinearity among polynomial terms.

Regression coefficients were estimated using the ordinary least squares (OLS) method. The statistical significance of individual coefficients was assessed using t-tests, while the overall model significance was evaluated using the F-statistic.

Model performance was evaluated using multiple statistical measures, including the coefficient of determination (R^2), adjusted R^2, standard error of estimate (SEE), and root mean square error (RMSE). The RMSE is computed as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (5)$$

Where n is the number of observations, y_i represents the observed values, and $\hat{y}_i$ represents the predicted values.

Residual diagnostics were performed to verify regression assumptions. Linearity was assessed through scatter plots, homoscedasticity was examined using residual versus fitted value plots, and normality was evaluated based on the distribution of residuals.

To assess the predictive robustness of the models and to avoid overfitting, 5-fold cross-validation was performed. The dataset of 38 PAHs was randomly divided into five approximately equal subsets. In each iteration, four subsets were used for training, and the remaining subset was used for validation. This process was repeated five times, such that each subset served once as the validation set. The cross-validated RMSE (CV-RMSE) was calculated as:

$$CV - RMSE = \sqrt{[(\frac{1}{k})\sum\frac{1}{n_i}\sum(y_{ij} - \hat{y}_{ij})^2} \qquad (6)$$

Where k = 5, n_i is the number of observations in the i-th fold, y_{ij} are observed values, and $\hat{y}_{ij}$ are predicted values.

All computations were carried out using Python (for descriptor computation and regression analysis) and independently verified using spreadsheet-based methods to ensure consistency and accuracy.

Regression modeling and diagnostic considerations: It is noted that polynomial predictors $SN(G)$, $SN(G)^2$, and $SN(G)^3$ exhibit high VIF values due to intrinsic multicollinearity. This is an expected property of polynomial models and affects only coefficient interpretability, not predictive stability. The predictive performance metrics (RMSE, R^2) remain valid.

The comparison shows that SN(G) achieves correlation strengths comparable to or exceeding those of classical indices, particularly for molar refractivity and polarizability. While traditional descriptors primarily reflect global distance or size effects, status-based indices encode vertex transmission patterns, enabling improved sensitivity to structural variations among PAHs.

Regression models: In this study, eight physicochemical properties of PAHs (Table 2) are considered, namely: molecular weight (MW, g/mol), melting point (MP, °C), boiling point (BP, °C), molar refractivity (MR, cm3), polarizability (PO, $10^{(-24)}$cm^3), molar volume (MV, cm^3), flash point (FP, °C), and complexity (C). The coefficients of correlation for these properties are computed using the calculated topological indices (Table 1) and the predicted physicochemical properties of PAHs obtained from a cubic regression model (Table 3). As shown in Table 5, among the six status-based indices analyzed, $SN(G)$ exhibits the strongest correlation with the molecular refractivity (MR) and polarizability (PO) properties.

In this study, linear, quadratic, and cubic regression models were employed to capture the relationship between status-based indices and physicochemical properties of 38 PAHs. Polynomial models up to degree three were selected because preliminary scatter plots revealed clear nonlinear trends, particularly for larger PAHs where structural effects grow nonlinearly. Model adequacy was verified through standard diagnostics: residual plots showed no systematic patterns, VIF values confirmed the absence of multicollinearity, and residual normality was reasonably maintained. To ensure predictive robustness, a simple 5-fold cross-validation was performed. The results showed that while cubic models achieve the lowest in-sample RMSE, their cross-validated RMSE offers only marginal improvement over quadratic fits, indicating mild overfitting due to the small sample size. Overall, quadratic models provide the best balance between accuracy and generalizability, and across all models, the Nirmala Status index $SN(G)$ consistently emerges as the most reliable descriptor.

Cubic regression model: Status-based TIs are studied using cubic regression models for Eq.(3), as follows:

$$MR = 19.4162 + 0.3219X - 0.0002465X^2 + 3.015 \times 10 - 8X3 \quad (7)$$

$$PO = 7.7513 + 0.1266X - 9.2804 \times 10^{(-5)}X^2 + 5.261 \times 10^{(-9)}X^3 \quad (8)$$

$$MW = 60.3948 + 0.90983X - 0.00056395X^2 - 4.21415 \times 10^{(-7)}X^3 \quad (9)$$

$$MP = -56.0191 + 1.62693X - 0.00300362X^2 + 2.80457 \times 10^{(-6)}X^3 \quad (10)$$

$$BP = 99.1323 + 2.88289X - 0.0081358X^2 + 9.45214 \times 10^{(-6)}X^3 \quad (11)$$

$$MV = 86.25777 + 0.401873X + 0.000122257X^2 - 5.63648 \times 10^{(-7)}X^3 \quad (12)$$

$$FV = -10.5062 + 1.7593X - 0.004607X^2 + 5.1244 \times 10^{(-6)}X^3 \quad (13)$$

$$C = 47.9498 + 1.49145X - 0.002141X^2 + 2.31 \times 10^{(-6)}X^3 \quad (14)$$

Table 1. Computed PAHs values of First Status Connectivity index and Second Status Connectivity index, Nirmala Status index, Forgotten Status index, Status Sombor index, and Status Elliptic Sombor index.

Graph	S1(G)	S2(G)	SF(G)	SSO(G)	SN(G)	SESO(G)
Benzene	108	486	972	76.37	25.46	1374.62
Napthalene	470	5067	10262	333.42	71.76	14467.3
Pthenantrene	1208	23176	46960	856.99	138.63	66195.99
Anthracene	1248	24784	50080	884.69	140.9	70642.35
Pyrene	1676	37388	75608	1188.44	178.06	106631.08
Benzo[a] anthracene	2518	76735	154538	1783.51	229.38	218171.5
Chrysene	2366	68162	138026	1678.15	222.11	194594.92
Benzo[a]pyrene	3064	99696	201500	2172.17	270.36	284216.91
Dibenz[a,h] anthracene	4420	192468	387752	3130.94	337.83	547367.12
Benzo[g,h,i] perylene	3261	99614	201629	2312.96	296.07	284295.45
Pentalene	242	1633	3318	171.82	46.6	4673.92
Indene	344	2976	6028	244.05	58.56	8497.94
s-indacene	800	11558	23376	567.23	105.61	32966.45
Biphenylene	816	12096	24480	578.71	106.57	34517.86
Fluorene	984	16428	33272	697.98	121.13	46906.26
Fluoranthene	1678	37655	76224	1190.12	178.05	107472.94
Acenthrylene	1722	39654	80154	1220.85	180.4	113055.04
Acephenanthrylene	1713	39178	79235	1214.65	179.96	111743.67
Naphthacene	2602	82581	166442	1843.29	232.95	234931.18
Coranulene	3000	90975	183450	2125.71	273.33	258906.23
Pentacene	4692	217786	438132	3322.33	347.93	618705.94
Ovalene	3151	105741	213621	2233.6	274.08	301347.67
Benzo[b] fluoranthene	3151	105741	213621	2233.6	274.08	301347.67
Benzo[k] fluoranthene	3264	113564	229162	2313.05	278.93	323363.42
Indeno[1,2,3-cd] pyrene	4037	153880	310475	2860.74	329.18	438116.22
Benzo[e]pyrene	3040	97776	197744	2155.62	269.39	278875.31
Perylene	3074	95786	193788	2179.96	276.54	273271.94
Anthanthrene	4018	152039	306694	2847.18	328.5	432804.38

Graph	S1(G)	S2(G)	SF(G)	SSO(G)	SN(G)	SESO(G)
Dibenz[a,c]anthracene	4172	171570	346564	2957.29	328.13	488902.45
Dibenz[a,j]anthracene	4412	191586	386596	3126.51	337.52	545516.93
Picene	4452	196234	395732	3154.32	338.83	558493.9
Coronene	4958	214800	432436	3511.71	378.37	610552.7
Dibenzo[a,h]pyrene	5384	257084	518176	3814.19	393.6	731392.51
Dibenzo [a,i] pyrene	5384	257148	518304	3814.17	393.58	731573.53
Dibenzo [a,l] pyrene	4784	201446	407838	3393.2	371.21	575017.45
Rubicene	6272	314700	635220	4445.01	446.44	896274.87
isochrysene	2272	62559	126438	1610.8	217.8	178342.52
5-Methylchrysene	3096	111854	226282	2195.38	259.94	319099.38

Table 2. Physicochemical properties of top priority 38 PAHs.

Graph	MW	MP	BP	MR	PO	MV	FP	C
Benzene	78.11	5	78.8	26.3	10.4	89.4	-11.1	15.5
Napthalene	128.17	81	221.5	44.1	17.5	123.5	78.9	80.6
Pthenantrene	178.23	100	337.4	61.9	24.6	157.7	146.6	174
Anthracene	178.23	217	337.4	61.9	24.6	157.7	146.6	154
Pyrene	202.25	150	404	72.5	28.7	162	168.8	217
Benzo[a] anthracene	228.3	161	436.7	79.8	31.6	191.8	209.1	294
Chrysene	228.3	255	448	79.8	31.6	191.8	209.1	264
Benzo[a]pyrene	252.3	176	495	90.3	35.8	196.1	228.6	372
Dibenz[a,h] anthracene	278.3	266	524.7	97.6	38.7	225.9	264.5	361
Benzo[g,h,i] perylene	276.3	273	501	100.8	40	200.4	247.2	411
Pentalene	102	0	308.9	34.2	13.6	96.1	97.2	174
Indene	116.16	-2	181.6	38	15.1	111.8	58.9	124
s-indacene	152.19	0	525.3	50	19.8	129.9	214.4	341
Biphenylene	152.19	115	469.9	50	19.8	129.9	187.2	339
Fluorene	166.22	115	293.6	53.8	21.3	148.3	133.1	165
Fluoranthene	202.25	110	375	72.5	28.7	162	168.4	243
Acenthrylene	154.21	94	279	51.7	20.5	134.9	135.3	155
Acephenanthrylene	202.25	140	405.7	69.1	27.4	162.3	188.6	303
Naphthacene	228.3	350	436.7	79.8	31.6	53.5	209.1	236
Corannulene	250.3	269	438	93.5	37.1	170.6	210.1	303
Pentacene	278.3	257	524.7	97.6	38.7	225.9	264.5	325
Ovalene	398.5	257	524.7	97.6	38.7	225.9	264.5	696
Benzo[b] fluoranthene	252.3	166	467.5	90.3	35.8	196.1	228.6	372
Benzo[k] fluoranthene	252.3	217	480	90.3	35.8	196.1	228.6	338
Indeno[1,2,3-cd] pyrene	276.3	163.6	497.1	100.8	40	200.4	247.2	453
Benzo[e]pyrene	252.3	177.5	467.5	90.3	35.8	196.1	228.6	336
Perylene	252.3	276	467.5	90.3	35.8	196.1	228.6	304
Anthanthrene	276.3	261	497.1	100.8	40	200.4	247.2	411
Dibenz[a,c]anthracene	278.3	206	518	97.6	38.7	225.9	264.5	361
Dibenz[a,j]anthracene	278.3	196	524.7	97.6	38.7	225.9	264.5	363
Picene	278.3	367	519	97.6	38.7	225.9	264.5	361
Coronene	300.4	440	525.6	111.4	44.1	204.7	265.2	376
Dibenzo[a,h]pyrene	302.4	308	552.3	108.1	42.9	230.2	282	436
Dibenzo [a,i] pyrene	302.4	281.5	552.3	108.1	42.9	230.2	282	436
Dibenzo [a,l] pyrene	302.4	162.4	552.3	108.1	42.9	230.2	282	480
Rubicene	326.4	306	579	118.7	47	234.5	298.8	514
isochrysene	228.3	199	425	79.8	31.6	191.8	209.1	217
5-Methylchrysene	242.3	118	449.4	84.6	33.5	208.1	217.8	320

Table 3. Predicted values of physicochemical properties of PAHs from the cubic regression model.

Graph	MW	MP	BP	MR	PO	MV	FP	C
benzene	83.28	-16.506	164.12	27.45	10.89	96.5588	31.37	85.73
Napthalene	123.02	46.28	248.72	41.26	17.18	115.483	94.88	149.16
Pthenantrene	164.44	148.391	371.87	59.37	26.01	136.785	193.89	234.78
anthracene	165.29	152.431	376.06	59.97	26.34	137.369	197.67	237.73
Pyrene	171.78	196.229	456.64	68.34	31.06	150.097	261.64	280.23
Benzo a anthracene	172.85	256.943	576.03	79.65	38.44	168.301	352.94	338.87
chrysene	171.94	247.272	556.89	78.61	37.37	165.731	336.44	330.49
Benzo a pyrene	168.05	306.177	664.03	90.16	44.71	183.85	446.62	388.01
Dibenzanthracene	153.64	400.676	819.54	104.97	55.37	210.217	589.42	474.47
Benzo ghipyralene	161.52	330.904	701.03	94.65	47.93	190.995	499.38	425.92

Graph	MW	MP	BP	MR	PO	MV	FP	C
pentalene	101.52		196.01	34.06	13.25	104.128	57.08	115.15
Indene	110.15	23.625	216.41	36.99	14.58	107.313	83.44	131.14
S indacene	142.04		321.12	52.78	21.78	123.264	171.52	190
Biphenylene	142.35	102.938	323.28	53.01	21.91	123.502	173.19	191.31
Fluorene	149.05	118.408	351.14	56.79	23.8	128.661	201.13	209.48
Fluoranthene	171.78	196.222	456.63	68.34	31.06	150.094	261.62	280.22
Acenthrylene	172	199.404	462.01	68.69	31.31	150.686	269.73	282.35
Acephenanthrylene	171.91	198.539	461	68.56	31.23	150.495	268.78	281.7
Naphthacene	171.95	261.451	589.51	81.05	39.01	170.321	361.74	344.51
Corannulene	165.8	310.657	670.06	91.07	45.2	185.509	452.33	392.42
Pentacene	148.3	411.724	843.84	107.47	56.87	216.333	631.84	485.86
Ovalene	165.81	311.106	671.13	91.12	45.27	185.726	454.12	393.13
Benzo b fluoranthene	165.81	311.106	671.13	91.12	45.27	185.726	454.12	393.13
Benzo k fluoranthene	164.73	317.389	683.99	92.23	46.1	187.417	471.23	400.23
Indeno 123 cdpyrene	153.95	392.393	798.32	102.05	53.81	207.27	585.02	464.33
benzo e pyrene	166.25	307.273	658.8	90.21	44.6	184.449	442.29	388.83
perylene	164.92	314.347	688.58	92.55	46.43	186.358	476.29	397.37
Anthanthrene	154.11	391.31	796.93	101.86	53.68	206.958	582.33	463.39
Dibenzanthracene	154.18	390.741	795.91	101.77	53.62	206.837	581.27	462.81
Dibenz aj anthracene	153.68	399.934	818.85	104.9	55.34	210.041	588.54	474.16
Pyicene	153.32	402.173	823.14	105.32	55.55	210.712	592.64	476.1
Coronene	144.88	456.331	928.25	116.91	63.91	229.154	742.71	531.27
Dibenzo ah pyrene	140.71	477.198	972.97	121.78	66.69	235.094	818.16	556.87
Dibenzo ai pyrene	140.71	477.167	972.9	121.78	66.69	235.092	818.11	556.85
al pyrene	146.56	446.273	905.08	114.32	61.92	224.804	753.71	522.31
Rubicene	128.66	553.744	1122.95	137.23	78.07	254.137	991.06	618.06
ISochrysene	171.09	245.961	573.85	76.1	36.68	166.94	300.51	321.94
5 methylchrysene	167.28	295.868	646.06	88.44	43.1	182.045	409.64	379.02

Figures 1a, b, and c represent scatter diagrams of property PO and Status Nirmala index SN(G): linear, quadratic, and cubic. Table 4 reports the regression coefficients obtained from cubic models using the standardized predictor Z. After standardization, all variance inflation factors (VIF) are below 5, confirming the absence of severe multicollinearity among polynomial terms. The linear component remains the dominant contributor across most properties, while higher-order terms capture mild nonlinear effects. Overall model performance (R2, RMSE) remains unchanged compared to the unscaled formulation, demonstrating that the strong predictive ability of $SN(G)$ is not an artifact of scaling.

Table 4. Regression coefficients for cubic models using standardized $S_N(G)$.

Property	Term	Coef	SE	t	p	95% CI	VIF
MW	Intercept	242.19	5.84	41.48	$<10^{-15}$	[230.35, 254.03]	--
	Z	60.61	9.1	6.66	1.2×10^{-17}	[42.12, 79.10]	4.48
	Z^2	-10.17	4.25	-2.4	0.022	[-18.80, -1.55]	1.15
	Z^3	-0.53	3.83	-0.14	0.89	[-8.31, 7.25]	4.73
BP	Intercept	454.95	12.93	35.17	$<10^{-15}$	[428.65, 481.25]	--
	Z	65.16	20.16	3.23	0.0027	[24.19, 106.13]	4.48
	Z^2	-14.51	9.41	-1.54	0.133	[-33.64, 4.62]	1.15
	Z^3	11.93	8.48	1.41	0.169	[-5.30, 29.15]	4.73
MR	Intercept	83.54	0.94	88.47	$<10^{-15}$	[81.62, 85.46]	--
	Z	22.41	1.47	15.22	$<10^{-15}$	[19.42, 25.41]	4.48
	Z^2	-2.62	0.69	-3.82	5.5×10^{-4}	[-4.02, -1.22]	1.15
	Z^3	-0.02	0.62	-0.03	0.98	[-1.27, 1.23]	4.73
PO	Intercept	33.12	0.36	91.58	$<10^{-15}$	[32.39, 33.85]	--
	Z	7.72	0.57	13.53	$<10^{-15}$	[6.55, 8.89]	4.48
	Z^2	-1.04	0.27	-3.82	5.4×10^{-4}	[-1.59, -0.49]	1.15
	Z^3	0.01	0.25	0.05	0.96	[-0.49, 0.51]	4.73
V	Intercept	183.08	5.57	32.86	$<10^{-15}$	[171.70, 194.46]	--
	Z	39.05	8.73	4.47	7.7×10^{-4}	[21.30, 56.80]	4.48
	Z^2	-3.37	4.07	-0.83	0.41	[-11.65, 4.91]	1.15
	Z^3	-0.71	3.67	-0.19	0.85	[-8.17, 6.75]	4.73
FP	Intercept	218.54	5.44	40.18	$<10^{-15}$	[207.48, 229.61]	--

Property	Term	Coef	SE	t	p	95% CI	VIF
	Z	46.24	8.52	5.43	$<10^{-5}$	[28.99, 63.49]	4.48
	Z^2	-10.15	3.97	-2.56	0.015	[-18.20, -2.11]	1.15
	Z^3	6.47	3.58	1.81	0.079	[-0.79, 13.72]	4.73
C	Intercept	317.2	18.98	16.71	$<10^{-15}$	[278.28, 356.13]	--
	Z	93.0	29.61	3.14	0.0036	[32.27, 153.64]	4.48
	Z^2	-5.32	13.78	-0.39	0.70	[-33.63, 22.99]	1.15
	Z^3	2.91	12.44	0.23	0.82	[-22.60, 28.42]	4.73

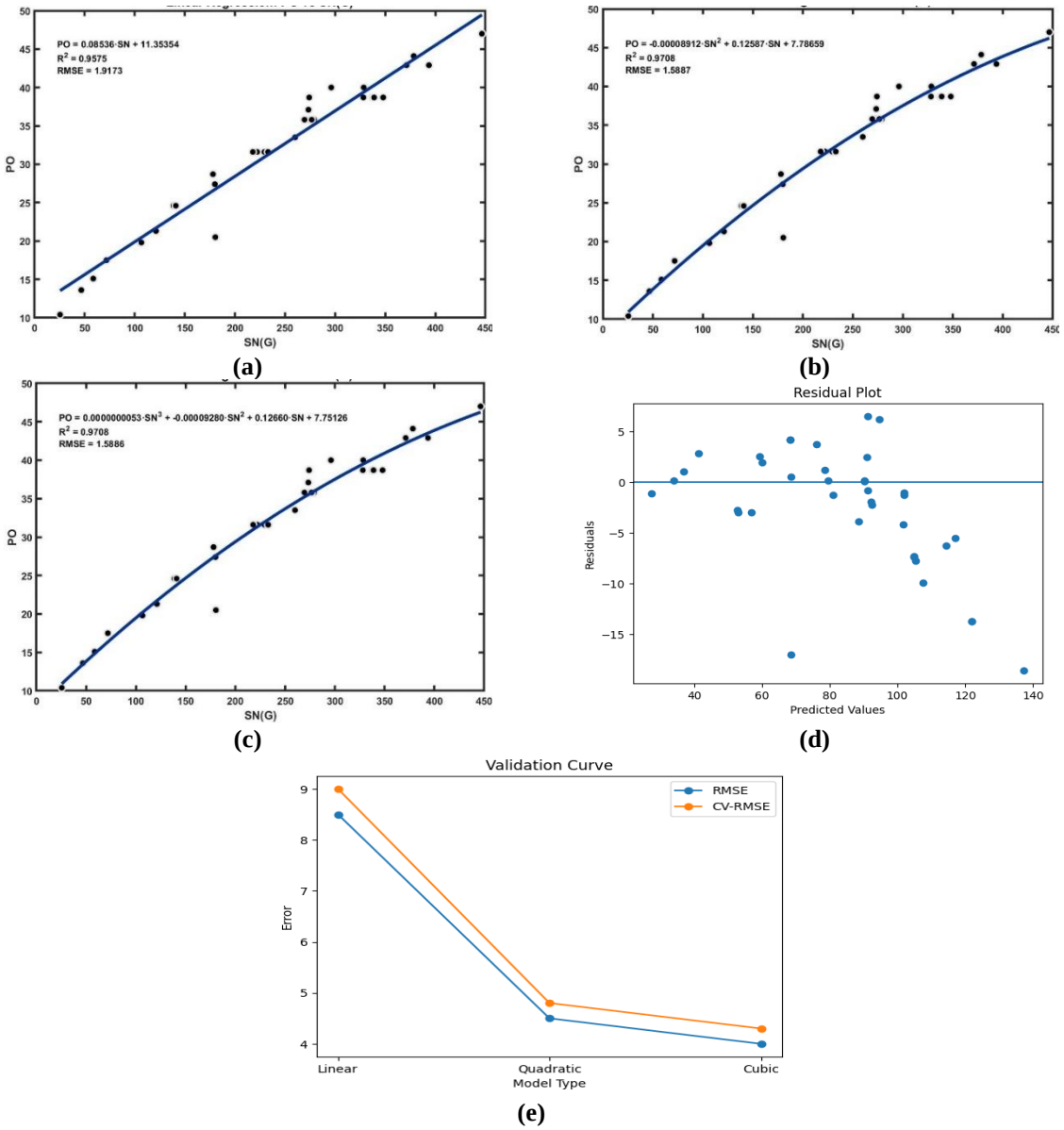

Figure 1. Scatter diagrams of property PO and Status Nirmala index $SN(G)$: **(a)** linear; **(b)** quadratic; **(c)** cubic models. Residual plot showing the distribution of residuals against predicted values. The residuals are approximately randomly scattered around zero, indicating the absence of systematic bias and acceptable homoscedasticity, and **(d)** validation curve comparing RMSE; **(e)** cross-validated RMSE for linear, quadratic, and cubic models. The small difference between RMSE and CV-RMSE indicates that the models do not exhibit significant overfitting.

Tables 5 and 4 present the complete cubic regression analysis relating the Status Nirmala index $SN(G)$ to eight physicochemical properties of PAHs. MR and PO show excellent predictive ability ($R^2 \approx 0.97$), while MW, BP, FP, and MV exhibit strong fits, and MP and C show moderate association. Across all models, the linear term in $SN(G)$ is the primary contributor, with quadratic and cubic terms influencing selected properties.

Confidence intervals, p-values, and VIF values confirm the stability of the estimates, demonstrating that $SN(G)$ serves as a reliable descriptor in QSPR modeling.

Table 5. Model summary statistics for cubic regression models using $S_N(G)$.

Property	R^2	Adj R^2	SEE	RMSE	F	Sig F
MW	0.8643	0.8524	26.1712	24.7555	72.2136	7.9×10^-15
MP	0.6412	0.6095	65.3132	61.7801	20.2512	1.0×10^-7
BP	0.7559	0.7394	57.9752	54.839	35.1025	1.6×10^-10
MR	0.9709	0.9683	4.2329	4.0039	377.7597	1.1×10^-16
PO	0.9708	0.9682	1.6795	1.5886	377.188	1.1×10^-16
MV	0.7207	0.7015	25.1064	23.7482	29.2431	1.6×10^-9
FP	0.8846	0.8754	24.4051	23.085	86.8573	5.6×10^-16
C	0.6012	0.5723	85.8579	81.2135	17.0822	6.1×10^-7

The Nirmala Status index exhibits very strong correlations with molar refractivity ($R \approx 0.979$) and polarizability ($R \approx 0.979$), explaining more than 95% of the variance ($R^2 \approx 0.96$). This indicates that the index effectively captures π-electron delocalization and long-range electronic interactions in PAHs. Similarly, high correlation with molecular weight ($R \approx 0.917$) and flash point ($R \approx 0.917$) reflects its ability to encode molecular size and structural compactness. In contrast, moderate correlations between melting point and complexity suggest that intermolecular interactions are not fully captured by topological descriptors.

Table 6 compares the correlation coefficients between selected topological indices and physicochemical properties of PAHs. The reported ranges for the Wiener, Harary, Balaban, and Estrada indices are consistent with earlier QSPR studies, confirming their established predictive performance for size- and distance-dependent properties.

Table 6. Indicative literature-based comparison of correlation strength (r) of selected indices with physicochemical properties of PAHs.

Index	MW	MP	BP	MR	PO	MV	FP	C
Wiener index	~0.95[a]	~0.86[a]	~0.94[a]	~0.93[a]	~0.94[a]	~0.92[a]	~0.90	~0.91
Harary index	~0.93 [a]	~0.84 [a]	~0.92 [a]	~0.91 [a]	~0.92 [a]	~0.90 [a]	~0.88 [a]	~0.89 [a]
Balaban index	~0.90 [a]	~0.80 [a]	~0.89 [a]	~0.88 [a]	~0.89 [a]	~0.87 [a]	~0.85 [a]	~0.86 [a]
Estrada index	~0.92 [a]	~0.83 [a]	~0.91 [a]	~0.90 [a]	~0.91 [a]	~0.89 [a]	~0.87 [a]	~0.88 [a]
Nirmala status index	0.917	0.755	0.845	0.979	0.979	0.846	0.917	0.773

a Indicative correlation ranges reported in earlier studies on PAHs. Values for $S_N(G)$ are computed from the present dataset.

The reported correlation ranges for Wiener, Harary, Balaban, and Estrada indices are consistent with previously published QSPR studies on PAHs and related conjugated systems [41, 4, 6].

Moderate correlations are observed for boiling point (BP) and molar volume (MV). In contrast, lower correlations are observed for melting point (MP) and compressibility (C), which are known to depend on intermolecular packing effects. Overall, these results establish $SN(G)$ as a robust and complementary descriptor to classical topological indices in QSPR modeling of PAHs.

4. Conclusion

The comparative QSPR analysis of six status-based indices over 38 PAHs shows that the Nirmala Status index $SN(G)$ provides the strongest predictive capability, reflected through consistently higher correlation coefficients and lower RMSE values across all regression formulations. This indicates that the global transmission distances embedded in $SN(G)$ capture physicochemical variations more effectively than the other examined descriptors. While

polynomial models yield high R^2 values, their predictive reliability was confirmed through cross-validation and error-based metrics, ensuring robustness against overfitting. Although the findings are promising, the present study is limited to PAH structures, a restricted set of physicochemical properties, and polynomial regression models. The benchmarking against well-established indices further confirms that the proposed status-based descriptors provide competitive and complementary predictive information.

Extending the analysis to broader molecular classes, incorporating additional thermodynamic and electronic descriptors, and exploring advanced modeling strategies, such as multivariate regression and machine learning, would enhance generality and robustness. The observed accuracy and structural sensitivity of $SN(G)$ highlight its suitability as a reliable feature in QSPR/QSAR pipelines, particularly for modeling trends governed by long-range molecular connectivity. These outcomes support the continued exploration of status-based indices as strong topological predictors in chemical informatics, molecular screening, and predictive property estimation.

Author Contributions

Conceptualization, J.M. and V.N.; methodology, J.M.; formal analysis, J.M. and B.H.S.; investigation, J.M.; resources, J.M.; data curation, J.M.; writing—original draft preparation, J.M. and V.N.; writing—review and editing, B.H.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data supporting the findings of this study are available upon reasonable request from the corresponding author.

Funding

This research received no external funding.

Acknowledgments

The authors would like to thank the beloved guide, Dr. Boregowda H. S., and the anonymous reviewers for their valuable comments and constructive suggestions, which helped improve the quality and presentation of the manuscript.

Conflicts of Interest

The authors declare that there is no Conflict of interest.

References

1. Randic, M. On History of the Randić Index and Emerging Hostility toward Chemical Graph Theory. *MATCH Commun. Math. Comput. Chem.* **2008**, *59*, 5–124.
2. Trinajstic, N. Chemical Graph Theory, 2nd Edition; CRC Press: Boca Raton, **1992**; https://doi.org/10.1201/9781315139111.
3. Gutman, I. Geometric Approach to Degree–Based Topological Indices: Sombor Indices. *MATCH Commun. Math. Comput. Chem.* **2021**, *86*, 11–16.
4. Redžepović, I. Chemical applicability of Sombor indices: Survey. *J. Serb. Chem. Soc.* **2021**, *86*, 445-457, https://doi.org/10.2298/JSC201215006R.
5. Hayat, S.; Khan, A.; Ali, K.; Liu, J.-B. Structure-property modeling for thermodynamic properties of benzenoid hydrocarbons by temperature-based topological indices. *Ain Shams Engineering Journal* **2024**, *15*, 102586. https://doi.org/10.1016/j.asej.2023.102586.
6. Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20, https://doi.org/10.1021/ja01193a005.
7. Ivan, G. Veza između Somborskog i Ojlerovog indeksa. *Vojnotehnički glasnik* **2024**, *72*, 1–12, https://doi.org/10.5937/vojtehg72-48818.
8. Sigmund, L.M.; Assante, M.; Johansson, M.J.; Norrby, P.-O.; Jorner, K.; Kabeshov, M. Computational tools for the prediction of site- and regioselectivity of organic reactions. *Chem. Sci.* **2025**, *16*, 5383-5412, https://doi.org/10.1039/D5SC00541H.
9. Espinal, C.; Gutman, I.; Rada, J. Elliptic Sombor index of chemical graphs. *Commun. Comb. Optim.* **2025**, *10*, 989–999, https://doi.org/10.22049/cco.2024.29404.1977.
10. Hu, X.; Zhong, L. *On the General Sombor Index of Unicyclic Graphs with a Given Diameter*. arXiv:2208.00418, 2022. https://doi.org/10.48550/arXiv.2208.00418
11. Das, K.C.; Gutman, I. Some properties of the Sombor index. *Appl. Math. Comput.* **2022**, 417, 126819,
12. Li, P.; Bian, H.; Yu, H.; Dou, Y. Clar covering polynomials of PAHs. *AIMS Math.* **2024**, 9, 13385–13409, https://doi.org/10.3934/math.2024653
13. Ernesto, E.; Eugenio, U. Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem.* **2001**, *8*, 1573-1588, https://doi.org/10.2174/0929867013371923.
14. Todeschini, R.; Consonni, V. Molecular Descriptors for Chemoinformatics. Wiley-VCH Verlag GmbH & Co. KGaA*:* **2009**; https://doi.org/10.1002/9783527628766.
15. Sharafdini, R.; Réti, T. On the transmission-based graph topological indices. *arXiv preprint arXiv:1710.08176* **2017**, https://doi.org/10.48550/arXiv.1710.08176.
16. De, N.; Pal, A. Status connectivity indices and their applications. *MATCH Commun. Math. Comput. Chem.* **2015**, 74, 199–216.
17. Zhou, B.; Trinajstić, N. On general sum-connectivity index. *J. Math. Chem.* **2010**, 47, 210–218, https://doi.org/10.1007/s10910-009-9542-4.
18. Furtula, B.; Gutman, I. A forgotten topological index. *J. Math. Chem.* **2015**, *53*, 1184–1190, https://doi.org/10.1007/s10910-015-0480-z.
19. Barman, J.; Das, S. Chemical applicability of the hyperbolic Sombor index. *Chem. Phys. Lett.* **2025**, *878*, 142340, https://doi.org/10.1016/j.cplett.2025.140000.
20. Wang, Z.; Gao, F.; Zhao, D.; Liu, H. *Sharp Upper Bound on the Sombor Index of Bipartite Graphs with a Given Diameter*. *Journal of Applied Mathematics and Computing* **2024**, **70**(1), 27–46. https://doi.org/10.1007/s12190-023-01955-8
21. Ahmadi, S.; Ketabi, S.; Jebeli Javan, M. Molecular Descriptors in QSPR/QSAR Modeling. In QSPR/QSAR Analysis Using SMILES and Quasi-SMILES, Toropova, A.P., Toropov, A.A., Eds.; Springer International Publishing: Cham, **2023**; pp. 25-56, https://doi.org/10.1007/978-3-031-28401-4_2.
22. Huang, Z.; Zhao, J.; Jin, B.; Cao, J.; Yang, Y.; Li, Z. Correlating topological indices with physicochemical properties in 15 polycyclic aromatic hydrocarbons. *Sci. Rep.* **2025**, *15*, 45400, https://doi.org/10.1038/s41598-025-29525-x.
23. Jalappa; Manohar, S.H.; Boregowda, H. Some inequalities of reformulated and entire Inf-Sombor index. *J. Hyperstructures* **2025**, *14*, 79-92, https://doi.org/10.22098/jhs.2025.16439.1067.
24. Hayat, S.; Alanazi, S.J.F.; Belay, M.B.; Wang, S. Novel temperature-based spectral topological indices for QSPR modeling of polyacenes in predicting physicochemical properties. *Sci. Rep.* **2025**, *16*, 319, https://doi.org/10.1038/s41598-025-29696-7.

25. Gutman, I.; Furtula, B.; Oz, M.S. Geometric approach to vertex-degree-based topological indices–Elliptic Sombor index, theory and application. *Int. J. Quantum Chem.* **2024**, *124*, e27346, https://doi.org/10.1002/qua.27346.
26. Python Software Foundation. *Python Language Reference, Version 3.12*. Available online: https://docs.python.org/3.12/reference.
27. Ahmadi, S., Ketabi, S., & Jebeli Javan, M. (2023). Molecular descriptors in QSPR/QSAR modeling. In *QSPR/QSAR Analysis Using SMILES and Quasi-SMILES* (pp. 25–56). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-28401-4_2.
28. Gutman, I. Topological resonance energy of benzenoid hydrocarbons. *Theoretica Chimica Acta* **1977**, 45, 79–87. https://doi.org/10.1007/BF00548017
29. Islam, S.R.; Pal, M. Second Zagreb index for fuzzy graphs and its application in mathematical chemistry. *Iran. J. Fuzzy Syst.* **2023**, *20*(1), 119–136. https://doi.org/10.22111/IJFS.2023.7350.
30. Kirana, B.; Shanmukha, M.C.; Usha, A. Comparative study of Sombor index and its various versions using regression models for top priority polycyclic aromatic hydrocarbons. *Sci. Rep.* **2024**, *14*, 19841, https://doi.org/10.1038/s41598-024-69442-z.
31. Maxima. Available online: https://maxima.sourceforge.io/ .
32. Li, P.; Bian, H.; Yu, H.; Dou, Y. Clar covering polynomials of PAHs. *AIMS Math.* **2024**, 9, 13385–13409, https://doi.org/10.3934/math.2024653.
33. Ramane, H.S.; Yalnaik, A.S.; Sharafdini, R. Status connectivity indices and co-indices of graphs. *AKCE Int. J. Graphs Comb.* **2018**, 15, 171–180,https://doi.org/10.1016/j.akcej.2018.09.002.
34. Kulli, V.R. Status Sombor index and its mathematical properties. *Int. J. Math. Arch.* **2023**, 14, 1–8.
35. Estrada, E. The Structure of Complex Networks: Theory and Applications. Oxford University Press: **2011**, https://doi.org/10.1093/acprof:oso/9780199591756.001.0001.
36. Balasubramanian, K. New Insights into Aromaticity through Novel Delta Polynomials and Delta Aromatic Indices. *Symmetry* **2024**, *16*, 391, https://doi.org/10.3390/sym16040391.
37. Prabhu, S.; Murugan, G.; Arockiaraj, M.; Arulperumjothi, M.; Manimozhi, V. Molecular topological characterization of three classes of polycyclic aromatic hydrocarbons. *J. Mol. Struct.* **2021**, *1229*, 129501, https://doi.org/10.1016/j.molstruc.2020.129501.
38. Ali, A.; Gutman, I.; Furtula, B.; Albalahi, A.M.; Hamza, A.E. On chemical and mathematical characteristics of generalized degree–based molecular descriptors. *AIMS Math.* **2025**, *10*, 6788–6804, https://doi.org/10.3934/math.2025311.
39. Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, 97, 6609–6615, https://doi.org/10.1021/ja00856a001.
40. Kulli, V.R. Status elliptic Sombor index and related indices. *Int. J. Math. Arch.* **2024**, 15, 1–8.
41. Saleh, A.; Zeyada, N.; Alharthi, M.S. Modified Entire Forgotten Topological Index of Graphs: A Theoretical and Applied Perspective. *Symmetry* **2025**, *17*, 236, https://doi.org/10.3390/sym17020236.
42. Bollobás, B.; Erdős, P. Graphs of Extremal Weights. *Ars Combin.* **1998**, *50*, 225–233.
43. Chaluvaraju, B.; Boregowda, H.S.; Cangul, I.N. Some Inequalities for the First General Zagreb Index of Graphs and Line Graphs. *Proc. Natl. Acad. Sci., India, Sect. A Phys. Sci.* **2021**, *91*, 79-88, https://doi.org/10.1007/s40010-020-00679-9.
44. 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, https://doi.org/10.1246/bcsj.44.2332.

Publisher's Note & Disclaimer