MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **96** (2026) 917–945
ISSN: 0340–6253
doi: 10.46793/match.96-3.20825

Relationship of Graph Energies with Some Graph Features

Hafsah Tabassum[a,b], **Nathakhun Wiroonsri**[a,b], **Pawaton Kaemawichanurat**[a,b,*]

[a] *Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok, Thailand*

[b] *Mathematics and Statistics with Applications (MaSA)*

hafsahtabassum@yahoo.com, nathakhun.wir@kmutt.ac.th, pawaton.kae@kmutt.ac.th

(Received August 26, 2025)

Abstract

Topological indices are widely used for identifying structure-property relationships due to their computational simplicity. Graph energies are an active research area nowadays, with over a thousand publications and an average of two papers published weekly. This study has two parts. The first part aims to explore the relationship between various graph energies of random tree graphs and well-known graph structural features, and the second part utilizes energies to predict the physicochemical properties of real-world molecular graphs. For both studies, we used XGBoost and SHAP (SHapley Additive Explanations) to build a decision-making model. We employed the Randomised Search CV to enhance XGBoost's performance further. This algorithm randomly selects a set of hyperparameters and evaluates the model's performance using cross-validation, resulting in improved accuracy. According to our research, XGBoost and SHAP (SHapley Additive Explanations) can help examine the relationships between topological indices and the structural features and physicochemical properties of drug molecules.

*Corresponding author.

1 Introduction

Many fields of chemistry require molecular descriptors to model QSPR and QSAR. Topological indices are popular for finding structure-property relationships due to their computational simplicity. There are hundreds of topological descriptors. The origin of their definition parameters can easily classify them. Degree, distance, and eigenvalue characterize topological molecular descriptors. This study examines topological descriptors based on eigenvalues. Molecular topological descriptors based on eigenvalues have been prominent since HMO theory showed their physical importance. This study emerged in the 1970s. However, eigenvalue-based descriptors have become so well-researched that they are now considered a branch of graph theory, graph spectral theory. Graph energies are matrix energies of various graph forms for each symmetric graph matrix. Quantitative chemistry calls them spectral indexes. Spectrum indexes might be single eigenvalues or matrix spectrum functions.

More than 100 graph energies have been defined using matrices other than the adjacency matrix [17, 18]. Graph invariants based on vertex degrees are documented in mathematical and chemical literature [11, 24]. The incidence and Sombor energies can be defined on the incidence and Sombor matrices, respectively. Similarly, Laplacian and Randić energies are defined over their corresponding matrices. A functional formula based on eigenvalues can quantify each of these energies. Matrix energy is usually the sum of the absolute eigenvalues of the simple adjacency matrix, but some other variants also exist.

Graph energies are an active study area nowadays, with over a thousand publications and an average of two papers per week (according to research data). The growth of graph energies is due to their unexpected applications in various engineering and science fields [7, 10], such as air transportation [23], face recognition [2], protein sequence comparison [12], high satellite resolution [1], spacecraft construction, crystallography [37], and complex networks. Other medical uses have been found. Stevanović, D. and Stanković [32] examined the association between simple and Laplacian energy versions in graphs. Mikołaj Morzy et al. discovered the corre-

lation between energy and centrality measurements in egocentric networks. Shao, Yanling, et al. investigated degree-based tree energies' upper and lower bounds [31]. Tabassum, H. et al. studied the relationship between Ordinary, Laplacian, Randić, Incidence, and Sombor Energies of Trees [22].

Among the many graph energy measures proposed in the literature, we focus on five fundamental types that, when combined, provide a comprehensive structural characterisation of graphs. Ordinary energy is a fundamental spectral descriptor that captures the overall eigenvalue distribution of the adjacency matrix [17]. Randić energy emphasises branching and chemical complexity through degree-based weighting [5], whereas Sombor energy considers graph geometry and distance [29]. Laplacian energy reveals the graph's connectedness and flow properties, as observed through the spectrum of the Laplacian matrix [21]. The incidence energy, calculated from the singular values of the incidence matrix, reveals patterns in edge-vertex interactions [19]. This carefully selected set, grounded in theoretical relevance and computational feasibility, aligns with the significant graph categories of energy measures recognized in the literature.

This study aims to learn more about the relationship of ordinary, Randic, Laplacian, Sombor and Incidence energies of random tree graphs with some well-known graph features, including maximum degree, average eccentricity, diameter, average shortest path length and mean, standard deviation, and minimum and maximum values of betweenness centrality, closeness centrality, and eigenvector centrality using machine learning algorithms. These traits were chosen because they encompass essential elements of graph topology—degree distribution, distance metrics, and centrality—that are expected to influence or correlate with the spectral properties contained in the graph's energy. We also studied the relationship between these five graph energies and the physicochemical properties of molecular graphs, including boiling point, density, enthalpy, flash Point, surface Tension, polarizability, log P, molar weight, molar volume, and molar refraction.

This article is structured as follows: Preliminaries, including the types of energy considered in this work, are defined in the next section. In Section 3, we give the methodology and computational details. All five

energies were analyzed for their relationship with fifteen features. Some findings and discussions are presented in Section 4. In Section 5, we explore the relationship between physicochemical properties and the five selected graph energies. Our findings are summarized and concluded in Section 6.

2 Preliminaries

Let $V(G)$ be a vertex set of an un-directed graph G and $E(G)$ be an edge set. Let n and m be the number of vertices and edges, respectively. If the vertices u and $v \in V(G)$ are adjacent, then uv denotes the edge between these vertices. Let d_u and d_v denote degree of the vertex u and v respectively. Graph energy is defined as the sum of the absolute value of eigenvalues of a graph G given by $E(G) = \sum_{j=1}^{n} |\lambda_j|$. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be its eigenvalues [10], for each $1 \leq j \leq n$, λ_j be the roots of the characteristic polynomial $\phi(G; x) = det(xI - A(G))$ where $A(G)$ represents the adjacency matrix of graph G. For the graph G, its adjacency matrix $A(G) = (a_{ij})_{n \times n}$ is a symmetric matrix of order n, whose elements are defined in [9] as:

$$a_{ij} = \begin{cases} 1 & \text{if } uv \in E(G) \\ 0 & \text{if } uv \notin E(G) \end{cases}$$

$$TI(G) = \sum_{uv \in E(G)} \phi(d_u, d_v)$$

where ϕ is suitable function with the condition $\phi(x, y) = \phi(y, x)$. The graph invariants stated above are known as topological indices. *Laplacian matrix* $L_{i,j}$ can be defined as:

$$L_{ij} = \begin{cases} d_v & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } uv \in E(G) \\ 0 & \text{otherwise} \end{cases}$$

Let λ_i for $1 \leq i \leq n$ be the eigenvalues of Laplacian matrix. The *Laplacian*

energy can be defined as:

$$LE(G) = \sum_{i=1}^{n} \left| \lambda_i - \frac{2m}{n} \right|$$

where m is cardinality of edge set and n be cardinality of vertex set of G. The Randić matrix [6] given by $R(G) = (r_{ij})_{n\times n}$ can be defined as:

$$r_{ij} = \begin{cases} \frac{1}{\sqrt{d_u d_v}} & \text{if } uv \in E(G) \\ \\ 0 & \text{if } uv \notin E(G) \end{cases}$$

The sum of absolute values of the eigenvalues of this Randić matrix is known as the Randić energy $RE(G)$ of the graph.

The incidence matrix $I(G)$ of an undirected graph G has a column for each edge and a row for each vertex.

$$I_{ij} = \begin{cases} 1 & \text{if vertex } v_i \text{ is incident to edge } e_j \\ 0 & otherwise \end{cases}$$

Incidence energy $IE(G)$ is the sum of the singular values of the incidence matrix $I(G)$ that are, in turn, equal to the square root of eigenvalues of $I(G)I(G)^t$ where $I(G)I(G)^t$ is a square matrix of order n.

The Sombor matrix, denoted by $A_{SO}(G) = (so_{ij})_{n\times n}$, of the graph G is a symmetric matrix having order n with the following elements:

$$so_{ij} = \begin{cases} \sqrt{d_u^2 + d_v^2} & \text{if } uv \in E(G) \\ \\ 0 & \text{if } uv \notin E(G) \end{cases}$$

The sum of the absolute value of eigenvalues of the Sombor matrix so_{ij} is known as the Sombor energy $SOE(G)$ of the graph.

A connected graph with no cyclic subgraph is a *tree*. The maximum degree of a tree is the degree of the vertex with the highest number of edges and is represented by $\Delta(T)$. The eccentricity of a vertex, denoted

as $e(V)$, is the greatest distance between that vertex and any other vertex. The diameter of a tree, denoted as $D(T)$, is defined as the longest distance between any two vertices in the tree. The path length of a tree is determined by adding up the lengths of all the paths from the root to each node in the tree. The average shortest path length from the root to a node is calculated by dividing the total length by the number of nodes in the tree. Several centrality measures for real-world networks and graphs have been introduced and investigated. The vertex properties are considered to rank them in terms of their relative importance inside the graph or network. Betweenness centrality is a concept that measures the extent to which a certain vertex is more central than all other vertices in a graph. Betweenness centrality quantifies the importance of a node in a graph by considering the shortest paths that pass through it. Within a tree, there exists a unique path connecting any two vertices. Hence, the betweenness centrality of a vertex v_i in a tree corresponds to the count of paths that traverse through that particular vertex. Mathematically, betweenness centrality for a tree can be written as

$$\mathcal{C}(n_1, n_2, \ldots, n_k) = \sum_{i<j} n_i n_j$$

The arguments n_i are the total number of vertices in the branches at v_i, eliminating v_i, arranged in any order. Closeness centrality, in the context of a connected graph, refers to a measure of centrality that quantifies the average distance between a specific vertex and all other vertices. Mathematically, it can be given as

$$\mathcal{C}_c(v_i) = \frac{1}{v} \sum_{v_j \in V} d_{ij}$$

The eigenvector centrality metric considers how many connections a vertex has (i.e., its degree) and the centrality of the vertices it is connected to. Mathematically, it can be given as

$$\mathcal{C}_E(v_i) = \frac{1}{\lambda} \sum_{v_j \in V_i} \mathcal{C}_E(v_j)$$

Computation of betweenness centrality, closeness centrality, and eigenvector centrality globally for the whole graph is computationally expensive. So, To compare betweenness centrality, closeness centrality, and eigenvector centrality with graph energies over a tree, we considered mean, standard deviation, maximum and minimum of betweenness, closeness and eigenvector centrality. Many classical graph parameters and centrality measurements for trees can be computed efficiently. Degree, diameter, eccentricity, average shortest path length (via the Wiener index), closeness centrality, and betweenness centrality are particularly conducive to linear-time algorithms, although eigenvector centrality can be computed to any established precision in time proportionate to edge number. The research carried out by Chindelevitch, Leonid, et al. on tree parameters and centralities summarizes these findings, providing precise $O(n)$ constraints for tree computations [8]. In contrast, the proper computation of graph energies such as ordinary, Laplacian, incidence, or related energies necessitates the whole set of eigenvalues or singular values of the associated adjacency- or Laplacian-based matrices. To acquire the whole spectrum, standard dense symmetric eigensolvers use tridiagonalization and QR iteration at a cost of $\frac{4}{3}n^3 + O(n^2)$ floating-point operations. The cubic complexity is described in *Matrix Computations* [15]. As a result, for big trees, exact centrality measures and classical graph parameters are significantly less expensive to compute than precise graph energies.

XGBoost, also known as *eXtreme Gradient Boosting*, is recognised as one of the most renowned machine learning algorithms. Gradient boosting is a machine learning methodology for many applications, such as classification and regression. The process involves creating a predictive model by combining several weak predictive models, usually decision trees. Gradient-boosted trees, as a strategy, generally outperform random forests when a decision tree is considered to be a weak learner. Like other boosting techniques, it constructs the model in stages but generalizes them by optimizing any differentiable loss function. The regularization strategy of XGBoost sets it apart from other gradient-boosting methods by avoiding overfitting.

We must tune a few hyperparameters affecting the model's predictions

while training an XGBoost model. Hyperparameter tuning is a powerful technique for enhancing the accuracy, precision, and other important features in supervised learning models. It involves searching for the best model parameters using various scoring techniques. To improve the model performance in XGBoost, we used a technique known as RandomizedSearchCV. RandomizedSearchCV provides the best set of hyperparameters by choosing combinations at random to produce the best score. By using this technique, we could fine-tune our model to obtain even better results. So, using RandomizedSearchCV along with XGBoost helped us to reach better regression accuracy for our dataset.

Cross-validation is a method that improves the precision of model prediction by splitting a data sample into two distinct subsets: the training set and the validation set. Multiple iterations of cross-validation utilize distinctive partitions, and the model's prediction accuracy is assessed by averaging the validation outcomes across these iterations.

A model's precision in machine learning is usually assessed by utilising the mean absolute percentage error (MAPE) and root mean squared error (RMSE). The Mean Absolute Percentage Error (MAPE) is a loss function used to quantify the magnitude of model error. The mean absolute percentage error (MAPE) is calculated by dividing the absolute difference between the actual and anticipated outcomes by the actual value. The ratios are aggregated for all values, and then the average is computed. More concisely, the formula for the MAPE is:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y_t} - y_t}{y_t}\right|.$$

where $\hat{y_t}$ is predicted value while y_t is actual value. *RMSE* i.e. Root Mean Square Error can be given as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y_t} - y_t)^2}.$$

Following the selection of the most effective model from XGBoost, we utilized SHAP (SHapley Additive Explanations) to evaluate the contribution

of each graph-based feature to the prediction of the energies under consideration. This interpretability analysis enhances our understanding of the model's decision-making process by revealing the relative relevance and directional influence of its features. Lundberg and Lee [25] introduced the Shapley Additive Explanation (SHAP), which has become popular for analysing machine learning model predictions. SHAP leverages Game Theory approaches [30] to understand how each feature contributes to specific predictions. This technique belongs to the additive feature attribution family and is model-agnostic, making it suitable for many machine learning and deep learning models. These methods help us understand the model's behaviour by assigning significance to specific input features. In the context of feature selection, SHAP-based approaches operate as follows: Classification models, such as XGBoost in this study, are trained on the entire dataset. SHAP values are then computed for each instance. For a model with a prediction function $g(x)$ and K features, we can obtain Shapley values as:

$$\phi_{\{i\}} = \sum_{P \subseteq N \setminus i} \frac{|P|!(K - |P| - 1)!}{M!} [g_x(S \cup \{i\}) - g_x(S)] \tag{1}$$

Where the number of feature value permutations that exist before the i-th feature value is denoted by $|P|!$. Likewise, $(|K|-|P|-1)!$ represents the total number of feature value permutations after the i-th feature value. The marginal contribution of adding the i-th feature value P is the difference term in the equation above. SHAP values are the solutions to Equation (1) under the assumptions: $g(x_p) = E[g(x|x_p)]$. That is, the prediction for any subset S of feature values is equal to the predicted value of the prediction for $g(x)$ given the subset x_p. These values are aggregated across the data set to determine the average absolute value of each feature. This method makes the computation of SHAP values more computationally demanding. The average SHAP value shows the average influence of each feature on model predictions across the dataset. In contrast, the absolute SHAP value reflects the feature's importance regardless of its direction (positive or negative). Sorting features by average absolute SHAP values in descending order identifies those with higher SHAP values as more

influential in the model predictions.

3 Methodology

This section includes the outcomes of the experimental analysis of relationships of five types of energies with some graph features. The trees considered in this work are all trees on n-vertices where $n = 6, 7, \ldots, 19$, that can be generated using the Python module Networkx.

Our data set consists of five energy measures and fifteen features: maximum degree, average eccentricity, diameter, average shortest path length and mean, standard deviation, and minimum and maximum values of betweenness centrality, closeness centrality, and eigenvector centrality, respectively. To provide a clearer overview, the methodological steps are summarized in the flow chart below.

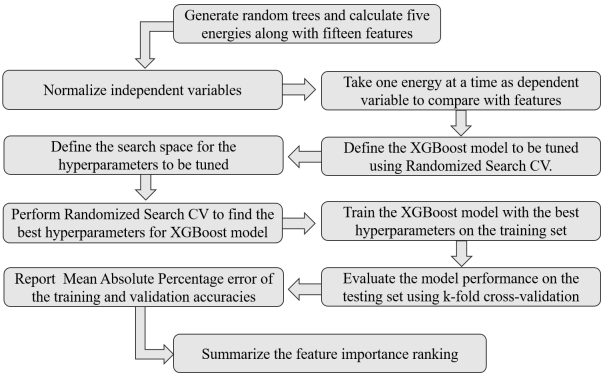

Figure 1. Flow Chart

4 Results and discussion

Using the Randomized Search CV, we discovered the following hyperparameters that improved the accuracy of the XGBoost method.
n_estimators: This parameter specifies the number of trees or rounds the model will try to learn. We set n-estimators between 1 to 1000. The best

estimators for smaller n, i.e., $6 \leq n \leq 10$, lie between 1 to 500, while for $11 \leq n \leq 19$, they lie between 800 and 1000 for our dataset.

max_depth: This parameter specifies the maximum depth of each tree in the model. For our dataset, we set max_depth between 2 to 8.

learning_ rate= 0.1: This parameter regulates the weight reduction of each model tree. The learning process will be slower, but the potential for increased accuracy is worth the wait. So, we set our learning rate between 0.1 to 0.5.

early_stopping_rounds: Parameter used for overfitting prevention. It allows the model to stop if there is no improvement in the learning.

n_jobs=-1: Allow to use all available CPU cores for faster search.

These parameters, when combined, determine the architecture of the XGBoost regressor model utilized in the project. The key component of our suggested dataset employing XGBoost and Randomized Search CV is the usage of k-fold cross-validation with 10-folds for $11 \leq n \leq 19$ and Leave-one-out for $6 \leq n \leq 10$. We may check the performance of our model on many subsets of data using k-fold cross-validation, lowering the risk of overfitting and boosting the robustness of our findings.

4.1 Feature importance via XGBoost

In the following tables maximum degree, average eccentricity, diameter, average shortest path length and mean, standard deviation, and minimum and maximum values of betweenness centrality, closeness centrality, and eigenvector centrality are represented by Δ, avg_ecc, D, l, BCA, BCM, BCD, BCM, CCA, CCD, $CCMin$, CCM, EV_CA, EV_CD, $EVCMin$ and $EVCM$ respectively. We are presenting five top most features that are important to predict the energies.

Table 1. Feature Ranking for Ordinary Energy

Vertices		Top 5 Features					RMSE	MAPE
	Feature Name	Δ	avg_ecc	D	l	BCA	0.6928	0.0805
6	importance	0.7819	0.2180	0	0	0		
	Feature Name	BCD	EV_CA	avg_ecc	EVCMin	Δ	0.5333	0.0641
7	importance	0.4387	0.3084	0.1255	0.0581	0.0558		
	Feature Name	BCD	avg_ecc	EVCM	Δ	EVCMin	0.4968	0.0459
8	importance	0.7397	0.0798	0.0465	0.0345	0.0290		
	Feature Name	BCD	avg_ecc	EVCM	Δ	EVCMin	0.4968	0.0459
9	importance	0.7397	0.0798	0.0465	0.0345	0.0290		
	Feature Name	BCD	Δ	avg_ecc	BCM	EVCMin	0.5232	0.0423
10	importance	0.6757	0.1197	0.0490	0.0377	0.0343		
	Feature Name	BCD	l	CCA	EVCM	Δ	0.4430	0.0317
11	importance	0.5737	0.1769	0.0750	0.0388	0.0266		
	Feature Name	BCD	Δ	avg_ecc	BCM	EVCM	0.3990	0.0255
12	importance	0.5408	0.2331	0.0420	0.0324	0.0309		
	Feature Name	BCD	Δ	l	avg_ecc	CCM	0.4060	0.0232
13	importance	0.4647	0.1762	0.0781	0.0717	0.0682		
	Feature Name	Δ	BCD	D	CCM	avg_ecc	0.4159	0.0222
14	importance	0.3281	0.3009	0.0846	0.0835	0.0536		
	Feature Name	Δ	BCD	D	avg_ecc	BCM	0.4216	0.0214
15	importance	0.6415	0.1001	0.0976	0.0668	0.0179		
	Feature Name	Δ	D	BCD	avg_ecc	BCM	0.3927	0.0180
16	importance	0.689	0.1193	0.0639	0.0331	0.0210		
	Feature Name	Δ	D	BCD	EVCM	BCM	0.4078	0.0178
17	importance	0.7237	0.1213	0.0374	0.0254	0.0173		
	Feature Name	Δ	D	EV_CD	BCD	EVCM	0.4053	0.0165
18	importance	0.6738	0.1398	0.0412	0.0329	0.0266		
	Feature Name	Δ	D	BCD	EVCM	EV_CD	0.4361	0.0168
19	importance	0.7068	0.1519	0.0299	0.0236	0.0146		

Table 2. Feature Ranking for Randic Energy

Vertices		Top 5 Features					RMSE	MAPE
	Feature Name	EV_CA	BCM	Δ	avg_ecc	D	0.3548	0.0644
6	importance	0.5125	0.4874	0	0	0		
	Feature Name	EVCMin	Δ	avg_ecc	D	l	0.4132	0.0800
7	importance	0.6876	0.3123	0	0	0		
	Feature Name	BCD	avg_ecc	Δ	EV_CA	EVCM	0.3839	0.0680
8	importance	0.4644	0.1491	0.1328	0.0841	0.0794		
	Feature Name	BCD	Δ	BCM	EVCMin	EV_CA	0.4073	0.0607
9	importance	0.6049	0.1275	0.0537	0.0475	0.0472		
	Feature Name	BCD	BCM	Δ	EVCMin	avg_ecc	0.4003	0.0597
10	importance	0.5522	0.0747	0.0670	0.0636	0.0571		
	Feature Name	BCD	Δ	BCM	CCMin	l	0.3460	0.0461
11	importance	0.5699	0.0727	0.0535	0.0446	0.0445		
	Feature Name	BCD	EV_CA	BCM	CCM	Δ	0.3653	0.0425
12	importance	0.592	0.0498	0.0494	0.0413	0.0379		
	Feature Name	BCD	avg_ecc	BCM	CCMin	Δ	0.3486	0.0377
13	importance	0.657	0.0693	0.0395	0.0388	0.0346		
	Feature Name	D	BCD	Δ	avg_ecc	BCM	0.3702	0.0361
14	importance	0.4822	0.2488	0.0675	0.0368	0.0268		
	Feature Name	D	BCD	Δ	avg_ecc	BCM	0.3776	0.0357
15	importance	0.3674	0.2124	0.1853	0.0459	0.0385		
	Feature Name	D	Δ	BCD	BCM	avg_ecc	0.3698	0.0322
16	importance	0.364	0.2651	0.1629	0.0373	0.0352		
	Feature Name	D	Δ	BCD	CCM	BCM	0.3791	0.0315
17	importance	0.3981	0.2360	0.1235	0.0630	0.0332		
	Feature Name	Δ	D	BCD	BCM	CCM	0.3820	0.0296
18	importance	0.4297	0.324	0.0748	0.027	0.0246		
	Feature Name	Δ	D	BCD	EVCM	BCM	0.4129	0.0305
19	importance	0.5478	0.2249	0.0610	0.0274	0.0252		

Table 1 shows that the maximum degree and standard deviation of betweenness centrality are the most important features to predict the Ordinary Energy of the graph, and the average of betweenness centrality has null importance for almost all the trees.

Table 2 shows that the standard deviation of betweenness, diameter, maximum degree, and the average of eigenvector centralities are the im-

portant features to predict the randic energy of the trees.

Table 3. Feature Ranking for Laplacian Energy

Vertices		Top 5 Features					RMSE	MAPE
	Feature Name	avg_ecc	Δ	BCM	D	1	4.6049	0.2100
6	importance	0.5721	0.3613	0.0664	0	0		
	Feature Name	Δ	avg_ecc	BCM	D	1	4.9883	0.1655
7	importance	0.6903	0.2347	0.0749	0	0		
	Feature Name	CCMin	Δ	BCD	avg_ecc	D	4.0176	0.1054
8	importance	0.6264	0.2655	0.1080	0	0		
	Feature Name	EVCM	BCD	1	Δ	BCM	4.2856	0.0699
9	importance	0.3906	0.3206	0.0633	0.0603	0.0586		
	Feature Name	BCD	BCM	CCMin	EV_CA	Δ	5.9038	0.0680
10	importance	0.4915	0.2931	0.0563	0.0399	0.0299		
	Feature Name	BCD	BCM	Δ	avg_ecc	CCMin	6.0416	0.0609
11	importance	0.4817	0.1552	0.0663	0.0627	0.0499		
	Feature Name	D	BCD	Δ	EV_CA	EVCMin	4.3689	0.0368
12	importance	0.2940	0.2891	0.0776	0.0575	0.0488		
	Feature Name	D	BCD	Δ	CCA	avg_ecc	4.7738	0.0362
13	importance	0.3169	0.2821	0.1130	0.0407	0.0347		
	Feature Name	Δ	BCD	D	1	EV_CD	5.7629	0.0335
14	importance	0.2807	0.2705	0.1710	0.0706	0.0325		
	Feature Name	Δ	BCD	D	1	BCM	4.8450	0.0291
15	importance	0.5948	0.0984	0.0923	0.0538	0.0374		
	Feature Name	Δ	D	BCD	1	EVCM	4.3769	0.0274
16	importance	0.6253	0.0924	0.0900	0.0431	0.0266		
	Feature Name	Δ	D	BCD	EVCM	BCM	4.6677	0.0235
17	importance	0.7131	0.0735	0.0549	0.0286	0.0263		
	Feature Name	Δ	D	BCD	BCM	EVCM	3.8564	0.0223
18	importance	0.7393	0.0731	0.0410	0.0273	0.0261		
	Feature Name	Δ	D	BCD	EVCM	BCM	3.9887	0.0219
19	importance	0.7704	0.0640	0.0381	0.0252	0.0225		

Table 4. Feature Ranking for Sombor Energy

Vertices		Top 5 Features					RMSE	MAPE
	Feature Name	Δ	BCM	avg_ecc	D	1	2.4504	0.1399
6	importance	0.9806	0.0193	0	0	0		
	Feature Name	Δ	avg_ecc	BCD	EV_CA	CCMin	2.0484	0.0878
7	importance	0.8950	0.0611	0.0294	0.0126	0.0017		
	Feature Name	Δ	CCMin	BCD	avg_ecc	BCM	1.4829	0.0420
8	importance	0.6118	0.2646	0.0680	0.0499	0.0054		
	Feature Name	EVCM	avg_ecc	1	BCD	Δ	2.0159	0.0490
9	importance	0.2322	0.2128	0.2100	0.1295	0.0925		
	Feature Name	Δ	BCD	1	avg_ecc	BCM	2.8214	0.0549
10	importance	0.2607	0.1800	0.1612	0.1013	0.0859		
	Feature Name	BCD	EVCM	CCD	BCM	avg_ecc	2.3123	0.0410
11	importance	0.2557	0.2169	0.0945	0.0901	0.0794		
	Feature Name	D	BCD	BCM	avg_ecc	CCMin	1.8183	0.0252
12	importance	0.3009	0.1839	0.1127	0.0579	0.0571		
	Feature Name	D	BCD	BCM	Δ	1	1.8240	0.0261
13	importance	0.3649	0.1636	0.0694	0.0646	0.0594		
	Feature Name	BCD	D	1	Δ	EV_CD	2.0680	0.0224
14	importance	0.2232	0.1809	0.1144	0.0752	0.0590		
	Feature Name	D	BCD	Δ	1	BCM	1.7455	0.0192
15	importance	0.3006	0.1542	0.1337	0.1146	0.0589		
	Feature Name	Δ	D	1	BCD	BCM	1.5630	0.0170
16	importance	0.3638	0.1737	0.1254	0.0989	0.0521		
	Feature Name	Δ	D	1	BCD	BCM	1.5409	0.0151
17	importance	0.4257	0.2170	0.0640	0.0590	0.0482		
	Feature Name	Δ	D	1	BCD	BCM	1.3990	0.0140
18	importance	0.5116	0.1578	0.0638	0.0589	0.0406		
	Feature Name	Δ	D	EV_CD	1	BCD	1.4402	0.0134
19	importance	0.5672	0.1039	0.0636	0.0568	0.0493		

Table 3 indicates that maximum degree, diameter, and standard deviation of betweenness centrality are the most important features to predict the Laplacian energy of trees. For trees on 6 and 7 vertices, average eccentricity is also important to predict Laplacian energy.

Table 4 indicates that the maximum degree and diameter are the most important features to predict the Sombor energy of trees. In the case of Sombor energy, the standard deviation of betweenness centrality seems least important compared to other energies. For trees on 9 and 11 vertices, maximum eigenvector centrality is also an important feature.

Table 5. Feature Ranking for Incidence Energy

Vertices		Top 5 Features					RMSE	MAPE
	Feature Name	Δ	avg_ecc	CCMin	D	l	0.0264	0.0031
6	importance	0.8477	0.1436	0.0086	0	0		
	Feature Name	avg_ecc	BCD	Δ	CCMin	BCM	0.0328	0.0042
7	importance	0.7374	0.1655	0.0701	0.0241	0.0027		
	Feature Name	l	CCA	BCD	avg_ecc	CCMin	0.0208	0.0017
8	importance	0.7904	0.1216	0.0371	0.0211	0.0149		
	Feature Name	l	BCD	avg_ecc	Δ	CCA	0.0077	0.0006
9	importance	0.5604	0.2412	0.1015	0.0551	0.0376		
	Feature Name	CCA	BCD	l	Δ	avg_ecc	0.0135	0.0007
10	importance	0.5588	0.2118	0.1155	0.0803	0.0255		
	Feature Name	l	CCA	BCD	Δ	EVCM	0.0105	0.0006
11	importance	0.4431	0.3582	0.1249	0.0561	0.0113		
	Feature Name	Δ	l	BCD	CCA	BCM	0.0104	0.0004
12	importance	0.5549	0.3382	0.0866	0.0121	0.0018		
	Feature Name	Δ	l	BCD	BCM	CCMin	0.0066	0.0003
13	importance	0.5831	0.3330	0.0734	0.0022	0.0017		
	Feature Name	l	Δ	BCD	CCA	BCM	0.0082	0.0003
14	importance	0.5373	0.4024	0.0410	0.0083	0.0020		
	Feature Name	Δ	l	BCD	CCA	EVCM	0.0069	0.0002
15	importance	0.5300	0.3914	0.0553	0.0068	0.0046		
	Feature Name	Δ	l	BCD	CCA	EVCM	0.0073	0.0002
16	importance	0.5129	0.4099	0.0479	0.0082	0.0057		
	Feature Name	Δ	l	D	BCD	CCM	0.0069	0.0002
17	importance	0.4971	0.3809	0.0528	0.0258	0.0149		
	Feature Name	Δ	l	D	BCD	CCM	0.0074	0.0002
18	importance	0.5541	0.3380	0.0297	0.0291	0.0146		
	Feature Name	Δ	l	BCD	D	CCM	0.0082	0.0002
19	importance	0.5620	0.3387	0.0283	0.0223	0.0134		

Table 5 shows that the maximum degree and average shortest path length are the most important features to predict the Incidence energy of trees. In the case of incidence energy, the trees' diameter seems least important. For trees on 7,10, and 11 vertices, average closeness centrality is also an important feature.

4.2 Feature importance via SHAP

To gain insights into the influence of each feature on the model's predictions, we employed SHAP (SHapley Additive exPlanations), a unified framework for interpreting complex machine learning models, as explained in the last section. The SHAP summary plots of random trees Corresponding to $N = 6, 7, \ldots, 19$ are given below. The goal is to determine which structural features have the most influence on each energy and how their values (low vs. high) affect the model's prediction.

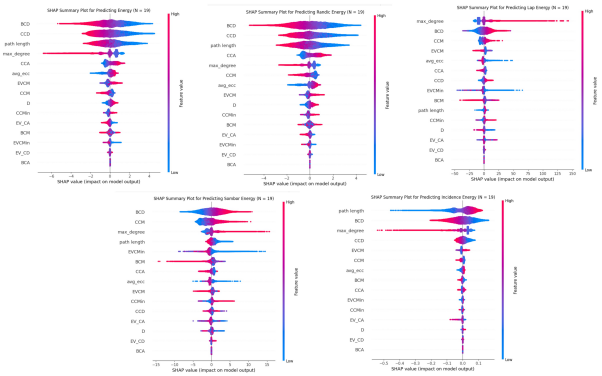

Figure 2. SHAP summary plots for energies on trees with $N = 19$ vertices.

Figure 2 demonstrates the SHAP summary plots for estimating Ordinary Energy (OE), Randic Energy (RE), Laplacian Energy (LE), Sombor Energy (SE), and Incidence Energy (IE). Each plot ranks features by importance (vertical axis), while the horizontal axis displays the SHAP value, which represents the impact of a feature on the relevant energy. Positive SHAP values indicate that the feature improves the expected energy, whilst negative values indicate a decreasing effect. The colour gradient represents the actual feature value, with red representing high and blue suggesting low values. This enables us to see not only what features are significant but also how their magnitude influences energy predictions. For example, in the case of Laplacian energy, the maximum degree appears as the most essential feature. High maximum degree values (in red) tend to increase the Laplacian energy, as evidenced by the concentration of red points on the right side of the SHAP axis, whereas lower values reduce it. For ordinary energy, the higher values of BCD, CCD, and average shortest path length decrease the prediction, while lower values increase it. Centrality measures of betweenness and closeness are also essential and follow the same trend as the maximum degree; higher BCD and CCM values increase the prediction, while lower values decrease it. Similarly, with Sombor Energy, metrics such as the standard deviation of betweenness centrality (BCD), the maximum of closeness centrality (CCM), and the average shortest path length follow the same pattern, with higher values

indicating increased energy.

In contrast, Incidence Energy exhibits low overall SHAP values across most features, indicating that it is relatively insensitive to structural variation in trees of this size. These findings show that graph energies respond differently to structural parameters, with ordinary, Randic, Laplacian, and Sombor energies being more susceptible to changes in centrality and connectedness, while incidence energy remains relatively stable. Given below are all the Plots of SHAP analysis plots for $N = 10$ and 15.

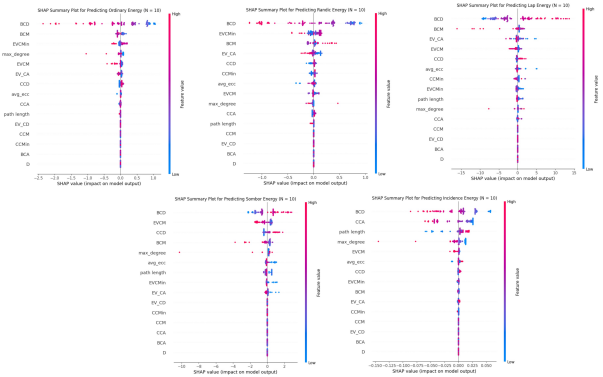

Figure 3. SHAP summary plots for energies on trees with $N = 10$ vertices.

Figure 3 shows in all models, betweenness-based measures (such as BCD and BCM) are always the most important predictors. Maximum degree is significant for ordinary, Sombor, and incidence energies. On the other hand, eigenvector- and closeness-based measurements are more critical for Laplacian and Randić energies. Average shortest path length and eccentricity show considerable importance, especially when it comes to incidence energy. Laplacian energy is the most sensitive to changes in features, whereas incidence energy is the least sensitive.

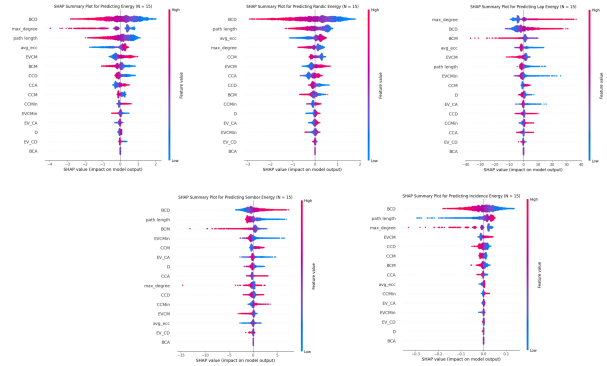

Figure 4. SHAP summary plots for energies on trees with $N = 15$ vertices.

The SHAP analysis for $N = 15$ (Figure 4) shows that betweenness-based descriptors (like BCD) and distance-related measurements (like average shortest path length and average eccentricity) are always the most important across all energy types. The maximum degree still has a significant effect, especially on ordinary and Laplacian energies. On the other hand, Sombor and incidence energies put more weight on path-based and betweenness metrics. The predictive models at $N = 15$ show more apparent feature importance separation than those at $N = 10$. BCD and average shortest path length stand out as strong universal predictors. Laplacian energy still shows the most sensitivity to features, whereas incidence energy is still not very sensitive. This indicates that different spectral energies have different dependence patterns. The next Section 4.3 explains the theoretical background of these dependence patterns.

4.3 Theoretical justification of feature importance

The results of the importance of the features derived in Sections 4.1,4.2 can be explained by the established principles of spectral graph theory. In this section, we point out critical mathematical relations that explain why some structural properties turned out to be the best predictors of graph energies. Throughout this discussion, stars and paths are frequently mentioned because they are the extreme instances for many spectral quantities:

among all trees with n vertices, stars typically maximize graph energies and eigenvalue bounds, whereas paths minimize them. General trees interpolate between these two extremes, resulting in the most important predictors being characteristics that separate star-like from path-like structures (such as degree distribution, standard deviation and betweenness, and diameter).

Maximum degree. The maximum degree $\Delta(G)$ is a common determinant of spectral behaviour across different energies. Because $\sqrt{\Delta(G)} \leq \lambda_1(G) \leq \Delta(G)$ [9] and $E(G) = \sum_i |\lambda_i|$ is dominated by λ_1, stars ($\Delta = n-1$) maximise energy while paths ($\Delta = 2$) minimise it [17]. For Laplacian and incidence energies, $\lambda_{\max}(L) \leq 2\Delta$ (Merris [26]), and since $II^T = L$, the same bound also governs incidence spectra. In the Randić energy scenario, degree heterogeneity causes spectral spread. Edges are weighted $1/\sqrt{d(u)d(v)}$, thus both large Δ and imbalance between high- and low-degree vertices enhance eigenvalues [6]. In case of the Sombor energy, edge weights scale as $\sqrt{d(u)^2 + d(v)^2}$, hence the effect is quadratic, making Δ disproportionately influential [20]. Although each energy has separate spectral definitions, their shared reliance on $\Delta(G)$ justifies its consistent top importance.

Betweenness centrality. Because trees have unique paths between each vertex pair, betweenness reflects the structural imbalance between stars and paths. The betweenness centrality of a vertex is defined as $BC(v)$ $= \#\{\text{vertex pairings whose unique path contains } v\}$. In ordinary, Laplacian, and incidence energies, extremal studies indicate that stars maximise spectra and paths minimise them. The variance or standard deviation of betweenness measures this differentiation. For Sombor energy, core hubs (high degree vertices) contribute edges with excessive $\sqrt{d(u)^2 + d(v)^2}$ weights, which are enhanced by high betweenness. For Randić energy, star-like topologies with a dominant hub shift the eigenvalue distribution compared to balanced path-like graphs, as most terms involve $1/\sqrt{\Delta \cdot d(v)}$. Thus, the distribution of betweenness values consistently indicates the tree design that produces energy extremes in all five definitions..

Diameter and eccentricity. Graph compactness, as assessed by diameter and eccentricity, is strongly related to spectral bounds. For Laplacian and incidence energies, Mohar's inequality $\lambda_{\max}(L) \geq n/\mathrm{diam}(G)$ [27] guarantees that small diameters deliver large eigenvalues. For ordinary energy, this translates into higher λ_1 and so higher energy. In Randić and Sombor energies, compact structures enhance spectral weights. In Randić, degree-normalized weights accumulate at the hub of stars (minimum diameter), while in Sombor, hubs with severe eccentricity values further amplify $\sqrt{d(u)^2 + d(v)^2}$. Thus, diameter and eccentricity are appropriate combined predictors across all energies.

Average Shortest path length. Average shortest path length is directly proportional to distance-based indices. For Laplacian energy, it connects to the Wiener index $W(G) = \sum_{\{u,v\}} d(u,v)$, which admits a Laplacian spectral decomposition (Merris [26]). Shorter pathways correlate with bigger Laplacian eigenvalues. Ordinary and incidence energies share this dependence since their spectra are likewise influenced by compactness. Graphs having a small diameter and a short average path length are considered more compact. Stars are the most compact trees (diameter $= 2$, average shortest path length around 2). Paths are the least compact (diameter $= n - 1$ and average shortest path length $\approx n/3$).Shorter paths in Randić and Sombor are associated with degree imbalance, resulting in more heterogeneity in $1/\sqrt{d(u)d(v)}$. In Sombor, path length reductions coincide with edge weights dominated by significant degrees. Thus, average shortest path length serves as a unifying distance-based explanation of energy fluctuation.

Other Centrality measures. The relevance of closeness centrality in ordinary, Laplacian, and incidence energies reflects the same compactness principle because it is inversely related to eccentricity and average shortest path length. Eigenvector centrality is formally defined as the Perron vector associated with the adjacency spectral radius λ_1, i.e., $A(G)x = \lambda_1 x$ with $x > 0$. It directly reflects ordinary energy and indirectly reflects Laplacian and incidence energies. While eigenvector centrality's predictive contri-

bution overlaps with $\Delta(G)$, it once again identifies hubs for Randić and Sombor. As a result, these centralities are theoretically consistent with the observed XGBoost rankings, although not as dominant as degree or path-based measures.

Conclusion. In conclusion, the importance feature rankings determined by XGBoost have been supported by solid findings from spectral graph theory. Centrality measurements strengthen degree- and distance-based dependencies, diameter and average shortest path length reflect compactness effects, and maximum degree and betweenness variance differentiate between star-like and path-like extremes. Tables and graphs in Sections 4.1, 4.2 show comparable patterns in feature importance, which can be explained by the simultaneous application of these structural principles to ordinary, Laplacian, incidence, Randić, and Sombor energies. After using XGBoost and SHAP to predict energies, we employed the same approach to predict the physicochemical properties of pharma- ceuticals and chemical compounds. The method involves using molecular graphs of alkanes (tree-like structures) to determine the five energies that were previously used. These energies are then used as descriptors for predicting physical and chemical properties, such as boiling and melting points, etc.

5 Molecular properties prediction of alkanes using graph energy descriptors

In theoretical chemistry, quantitative structure–property relationship (QSPR) investigations are crucial because they facilitate the estimation of the physicochemical and thermodynamic properties of molecular structures, particularly organic molecules. These predictive models utilize advanced mathematical and computational methodologies [3]. The concept of a "pathnumber," defined as the sum of pairwise distances, was initially introduced by Harold Wiener [36] to predict the boiling points of alkanes. In graph theory, this measure was later formalized as the Wiener index. In recent years, structure-based molecular descriptors have been employed

in QSPR modeling because they deliver the required mathematical frameworks. These invariants convert the molecular structure [16], omitting hydrogen atoms, into numerical values that capture significant chemical features. The evaluation of graphical invariants to determine their effectiveness in predicting physicochemical and thermodynamic properties has been a focal point of modern mathematical chemistry research. This technique filters out descriptors with lower predictive power while emphasizing those with enhanced predictive value.

Therefore, the goal of this section is to develop a quantitative structure-property-activity relationship (QSPR) between topological indices (energies) and specific physicochemical characteristics of alkanes to evaluate their effectiveness. We chose these compounds because of their noncyclic and tree-like structure. The list of alkanes included to predict physicochemical properties are given in Table 6 below.

Table 6. List of alkanes included in the study

Names of alkanes		
3-Methylpentane	2,3-Dimethyl-3-ethylhexane	2,4,4-Trimethylheptane
2-Methyl-3,3-diethylpentane	2,3,4,4-Tetramethylhexane	2,2-Dimethyl-4-ethylhexane
2,3,6-Trimethylheptane	3-Methyl-5-ethylheptane	2,3,3,5-Tetramethylhexane
2,2-Dimethyl-3-ethylhexane	2,3,5-Trimethylheptane	2-Methyl-4-ethylheptane
2,3,3,4-Tetramethylhexane	2,3,4,5-Tetramethylhexane	2,3,4-Trimethylheptane
2,2,4,4-Tetramethylhexane	2,2,5,5-Tetramethylhexane	3,4-Diethylhexane
2,3,3-Trimethylheptane	2,5-Dimethyl-3-ethylhexane	2,2,4,5-Tetramethylhexane
3,4,4-Trimethylheptane	2,2,6-Trimethylheptane	3,3,4-Trimethylheptane
2,2,3,5-Tetramethylhexane	3,3,5-Trimethylheptane	2,2,5-Trimethylheptane
4-Methyl-4-ethylheptane	2,2,3,4-Tetramethylhexane	2,5,5-Trimethylheptane
3,3-Diethylhexane	4-Methyl-3-ethylheptane	3,3-Dimethyl-4-ethylhexane
2,4,6-Trimethylheptane	3,3,4,4-Tetramethylhexane	3-Methyl-3-ethylheptane
2,3-Dimethyl-4-ethylhexane	2,4,5-Trimethylheptane	2,4-Dimethyl-3-isopropylpentane
2-Methyl-5-ethylheptane		

The methodology involves using molecular graphs of alkanes to calculate graph energy values. We represent each chemical structure of alkanes

as a graph and use adjacency, degree, and incidence matrices to figure out energies like ordinary, Randic, Sombor, Laplacian, and incidence energies. After that, these indices are combined with physicochemical properties, such as molecular weight, melting point, and boiling point, to create the final dataset. Next, the dataset is processed through an XGBoost-based pipeline that produces feature and target matrices, utilises randomised cross-validation to determine the optimal hyperparameters, and employs RMSE and MAPE to evaluate model performance. Lastly, SHAP values are calculated to help understand the model, indicating which features are essential and how easily the results can be interpreted.

5.1 Feature importance via XGBoost

The XGBoost algorithm was used to model the relationship between five graph energy-based features: Organizational Energy (E), Randic Energy (RE), Laplacian Energy (LE), Sombor Energy (SOE) and Incidence Energy (IE) and 7 molecular graph physicochemical properties, including boiling point (BP), flash point (FP), surface tension (ST), polarizability (Pol), log P (P), molar weight (MW), molar volume (MV), and molar refraction (MR). To improve the model performance, a hyperparameter tuning technique was created using RandomizedSearchCV, which efficiently explores the parameter space to find the optimal configuration. Following training, the model generated a preliminary ranking of feature importance using its gain metrics. This ranking, as shown in Table 7, revealed which energy descriptors were most important in making accurate predictions for each physicochemical property.

Table 7 shows how vital spectral energies are for predicting physicochemical qualities, as well as their error metrics. For the boiling point, ordinary energy OE is the most important predictor 0.712678, followed by Randić energy RE (0.227154), while other indices (LE, IE, SOE) only add a little bit. For surface tension, OE is again the most critical factor (0.560375), although IE and RE also play essential roles, followed by SOE and LE, which have less critical roles. OE (0.999152) is the only description that has a significant effect on polarisability. All other descriptors have minimal impact. Molar volume also depends almost entirely on OE

Table 7. Feature Ranking for Physicochemical Properties

Property	Feature Ranking					RMSE	MAPE
	1st	2nd	3rd	4th	5th		
Boiling Point	OE 0.712678	RE 0.227154	LE 0.036974	IE 0.020259	SOE 0.002936	2.512105	0.012415
Surface Tension	OE 0.560375	IE 0.215113	RE 0.132365	SOE 0.059423	LE 0.032724	0.535977	0.016189
Polarizability	OE 0.999152	IE 0.000262	RE 0.000209	LE 0.000188	SOE 0.000188	0.048603	0.001752
Log P	OE 0.760461	LE 0.239473	SOE 0.000048	IE 0.000018	RE 0.000000	0.099854	0.007165
Flash Point	OE 0.822142	RE 0.137374	LE 0.026266	IE 0.010013	SOE 0.004205	3.324495	0.068485
Molar Volume	OE 0.999029	SOE 0.000485	IE 0.000454	LE 0.000033	RE 0.000033	0.001312	0.323856
Molar Refraction	OE 0.96756	IE 0.0181	RE 0.005	SOE 0.004	LE 0.0039	3.0273	0.0331

(0.999029), with just minor effects from SOE, IE, and LE.

For Log P, OE is still the most critical factor 0.760461, followed by Laplacian energy LE (0.239473). SOE, IE, and RE have minimal effect. OE (0.822142) is the main factor that determines flash point prediction, followed by RE (0.137374). LE, IE, and SOE have only minor effects. OE (0.96756) is the main factor that affects molar refraction. IE (0.0181) and RE (0.005) also have minor effects, while SOE and LE don't matter much.

The error measures back up these ranks. The predictions for Log P $RMSE = (0.099854, MAPE = 0.007165)$, polarizability with ($RMSE =$ 0.048603) and ($MAPE = 0.001752$), Surface Tension ($RMSE =$ 0.535977), and ($MAPE = 0.016189$), Molar Refraction ($RMSE = 3.0273$, $MAPE = 0.0331$) and boiling point ($RMSE = 2.512105, MAPE =$ 0.012415) make good predictions. The flash point and the molecular volume, on the other hand, have the highest prediction error ($RMSE =$ $3.324495, MAPE = 0.068485$) and ($RMSE = 0.001312, MAPE =$ 0.323856), indicating that the model performs less effectively for this feature.

In general, the data show that OE is always the most critical descriptor for all physicochemical characteristics. This shows how important it is in spectral–property correlations. Other energies, including RE, IE, and LE, only affect specific properties, while SOE usually has a negligible effect. The fact that OE always wins shows that it is strong and dependable as a universal predictor for modelling physicochemical behaviours.

5.2 Feature importance via SHAP

SHapley Additive exPlanations (SHAP) were used in the second phase to gain a more in-depth and interpretable understanding of the model's decision-making process. SHAP values for each physicochemical property were calculated and averaged to produce an improved ranking of the five graph energy features. The plots in Figures 5 and 6 illustrate this SHAP-based ranking, which offers a clear perspective on the influence of features across all target properties.

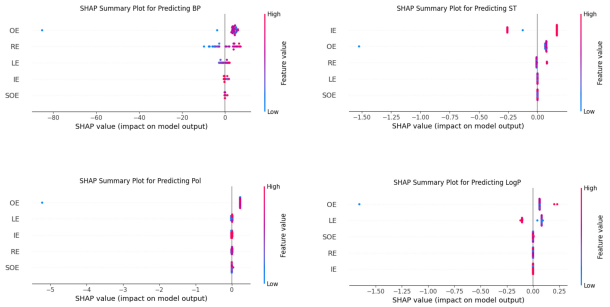

Figure 5. SHAP Summary Plot for Predicting Boiling Point, Surface Tension, Plarizability, Log P

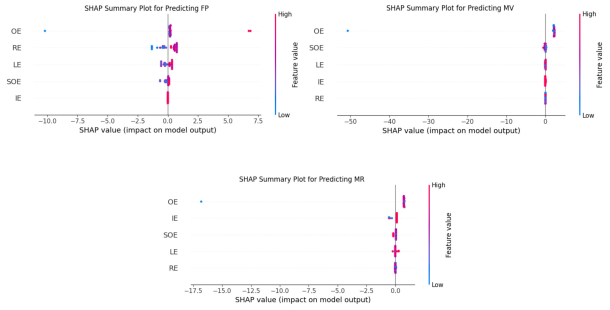

Figure 6. SHAP Summary Plot for Predicting Flash Point, Molar Volume, Molar Refraction

Figure 5 and 6 show SHAP summary charts that show how spectral energies affect the prediction of four physicochemical properties: boiling point (BP), surface tension (ST), polarisability (Pol), LogP (P), Flash Point (FP), Molar Volume(MV), and Molar Refraction (MR). Each plot

shows the SHAP values for each feature and ranks them by importance. The colour of the feature value (blue = low, red = high) represents the magnitude of the feature value. OE is the primary descriptor for boiling point, and large values always push the forecast up. RE has a smaller but still noticeable effect. The other indices (LE, IE, and SOE) have little to no effect. When it comes to surface tension, both OE and IE have a role, but OE is the more decisive factor. RE, LE, and SOE have effects that are not as strong as those of the other categories. OE explains practically all of polarizability, and the other descriptors don't add anything to it. This shows how much more important OE is than the others in feature ranking. LogP is mainly controlled by OE, with LE having a small but helpful role. SOE, RE, and IE are still on the edge of the distribution. The SHAP analysis backs up what we saw in Table 7 by showing that OE always has the most significant positive effect on model predictions, whereas other energies (RE, IE, and LE) only affect specific properties. This strengthens OE's position as the primary spectral determinant of physicochemical behaviour.

The SHAP summary plots for Flash Point (FP), Molar Volume (MV), and Molar Refraction (MR) show that overall energy (OE) is the most critical factor in making predictions. For FP, OE, and Randic energy (RE) are the most essential descriptors, while Laplacian energy (LE), Sombor energy (SOE), and incidence energy (IE) have minimal effect. For MV, OE alone has almost all of the prediction power, while other characteristics have minimal impact. For MR, OE remains the most critical factor, with only minor secondary effects from IE and SOE. LE and RE, on the other hand, are still not very important. These results show that ordinary energy OE is always necessary for varied physicochemical parameters, which strengthens its ability to predict.

6 Concluding remarks

Our study examines the characteristics of diverse graph energies within the framework of random trees. Graph energy refers to the energy of a symmetric matrix that depicts a tree topology. A variety of matrix types

can be used to represent a random tree. For example, adjacency, Randic, Laplacian, Sombor, and incidence matrices all give different energy values. When these energy values are used to describe the whole structure of a random tree, they don't provide much information, which could lead to confusion. Graph energy gives a vertex a unique description that goes along with topological properties like degree and distance-based traits like betweenness and average shortest path length. It is crucial to note that calculating centrality measures for the complete tree is computationally expensive, so we chose the standard deviation, maximum, minimum, and average of these values for ease of use. We observed a substantial association between graph energy and characteristics such as maximum degree, the standard deviation of betweenness centrality, and the diameter of random trees. The maximum degree is the most essential factor in estimating the energies of all the points since the energy of a graph is the sum of the energies of its vertices. This means that vertices with a higher degree add more energy to the tree.

In the second half, we used graph energies to predict properties such as boiling point, surface tension, polarizability, logP, flash point, molar volume, and molar refraction for alkanes. Our study, strengthened by feature ranking and SHAP values, consistently identified ordinary energy as the primary descriptor. In contrast, other energies such as Randic, Laplacian, incidence, and Sombor contributed in a property-specific manner, although of lesser significance. This shows that total energy is a flexible and dependable way to forecast molecule attributes, giving energy-based modelling both accuracy and interpretability.

***Acknowledgment*:** This research was supported by King Mongkut's University of Technology Thonburi's Postdoctoral Fellowship Under Research Project ID 28148. Pawaton Kaemawichanurat has been supported by National Research Council of Thailand (NRCT) and King Mongkut's University of Technology Thonburi (N42A660926).

References

[1] M. Akram, S. Naz, Energy of Pythagorean fuzzy graphs with applications, *Mathematics* **6** (2018) #136.

[2] A. S. Angadi, M. Hatture, Face recognition through symbolic modeling of face graphs and texture, *Intern. J. Pattern Recog. Artif. Intell.* **33** (2019) #1956008.

[3] S. C. Basak, D. Mills, Quantitative structure property relationships (QSPRS) for the estimation of vapor pressure: A hierarchical approach using mathematical structural descriptors, *J. Chem. Inf. Comput. Sci.* **41** (2001) 692–701.

[4] P. Bonacich, Power and centrality: A family of measures, *Am. J. Soc.* **92** (1987) 1170–1182.

[5] S. B. Bozkurt, D. Bozkurt, Randić energy and Randić Estrada index of a graph, *Eur. J. Pure Appl. Math.* **5** (2012) 88–99.

[6] S. B. Bozkurt, A. D. Güngör, I. Gutman, Randić matrix and Randić energy, *MATCH Commun. Math. Comput. Chem.* **64** (2010) 239–250.

[7] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton Univ. Press, Princeton, 2005.

[8] L. Chindelevitch, M. Hayati, A. F. Y. Poon, C. Colijn, Network science inspires novel tree shape statistics, *PLoS One* **16** (2021) #e0259877.

[9] D. Cvetković, P. Rowlinson, S. K. Simić, *An Introduction to the Theory of Graph Spectra*, Cambridge Univ. Press, Cambridge, 2010.

[10] D. Cvetković, M. Doob, H. Sachs, *Spectra of Graphs – Theory and Application*, Academic Press, New York, 1980.

[11] K. C. Das, I. Gutman, I. Milovanović, B. Furtula, Degree-based energies of graphs, *Lin. Algebra Appl.* **554** (2018) 185–204.

[12] L. Di Paola, G. Mei, A. Di Venere, A. Giuliani, Exploring the stability of dimers through protein structure topology, *Curr. Protein Peptide Sci.* **17** (2016) 30–36.

[13] L. C. Freeman, A set of measures of centrality based on betweenness, *Sociometry* **40** (1977) 35–41.

[14] L. C. Freeman, Centrality in social networks conceptual clarification, *Soc. Networks* **1** (1979) 215–239.

[15] G. H. Golub, C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, 2013.

[16] I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986.

[17] I. Gutman, The energy of a graph, *Ber. Math. Stat. Sekt. Forschungsz. Graz* **103** (1978) 1–22.

[18] I. Gutman, Total π-electron energy of benzenoid hydrocarbons, *Topics Curr. Chem.* **162** (1992) 29–63.

[19] I. Gutman, D. Kiani, M. Mirzakhah, On incidence energy of graphs, *MATCH Commun. Math. Comput. Chem.* **62** (2009) 573–580.

[20] I. Gutman, B. Furtula, K. C. Das, On Sombor index and its energy, *Appl. Math. Comput.* **399** (2021) #126018.

[21] I. Gutman, B. Zhou, Laplacian energy of a graph, *Lin. Algebra Appl.* **414** (2006) 29–37.

[22] H. Tabassum, P. Kaemawichanurat, Adeela, N. Wiroonsri, Relationship between ordinary, Laplacian, Randić, incidence, and Sombor energies of trees, *MATCH Commun. Math. Comput. Chem.* **90** (2023) 743–763.

[23] J. Jiang, R. Zhang, L. Guo, W. Li, X. Cai, Network aggregation process in multilayer air transportation networks, *Chin. Phys. Lett.* **33** (2016) #108901.

[24] X. Li, Z. Wang, Trees with the extremal spectral radius of weighted adjacency matrices among trees weighted by degree-based indices, *Lin. Algebra Appl.* **620** (2021) 61–75.

[25] S. M. Lundberg, S. I. Lee, A unified approach to interpreting model predictions, in: U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, R. Fergus (Eds.), *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Curran Assoc. Inc., Red Hook, 2017, pp. 4768–4777.

[26] R. Merris, Laplacian matrices of graphs: a survey, *Lin. Algebra Appl.* **197–198** (1994) 143–176.

[27] B. Mohar, The Laplacian spectrum of graphs, in: Y. Alavi, A. J. Schwenk (Eds.), *Graph Theory, Combinatorics, and Applications*, Wiley, New York, 1991, pp. 871–898.

[28] M. E. J. Newman, *Networks: An Introduction*, Oxford Univ. Press, Oxford, 2010.

[29] M. S. Reja, S. M. A. Nayeem, On Sombor index and graph energy of some chemically important graphs, *Ex. Counterex.* **6** (2024) #100158.

[30] L. Shapley, A value for n-person games, in: H. Kuhn, A. Tucker (Eds.), *Contributions to the Theory of Games II*, Princeton Univ. Press, Princeton, 1953, pp. 307–317.

[31] Y. Shao, Y. Gao, W. Gao, X. Zhao, Degree-based energies of trees, *Lin. Algebra Appl.* **621** (2021) 18–28.

[32] D. Stevanović, I. Stanković, M. Milošević, More on the relation between energy and Laplacian energy of graphs, *MATCH Commun. Math. Comput. Chem.* **61** (2009) 395–401.

[33] D. Stevanović, *Spectral Radius of Graphs*, Academic Press, San Diego, 2011.

[34] R. Tibshirani, T. Hastie, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, 2001.

[35] A. Ullah, Z. Shamsudin, S. Zaman, A. Hamraz, G. Saeedi, Network-based modeling of the molecular topology of fuchsine acid dye with respect to some irregular molecular descriptors, *J. Chem.* **2022** (2022) #8131276.

[36] H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.

[37] K. Yuge, Graph representation for configuration properties of crystalline solids, *J. Phys. Soc. Japan* **86** (2017) #024802.